1.

Write the conditional bias of the local polynomial regression estimator for $p - \nu$ odd

$$\text{bias}[\hat{m}_\nu(x_0) \mid \mathbb{X}] = \varepsilon_{\nu+1}^T \mathbf{S}^{-1} \mathbf{c}_p \frac{\nu!}{(p+1)!} m^{(p+1)}(x_0) h^{p+1-\nu} + o_p(h^{p+1-\nu})$$

in terms of the equivalent kernel $K^\star_{\nu,p}$ (see p. 60 Eq. (4.29) in the notes).

Answer

Let $(X_1, Y_1), \cdots (X_n, Y_n)$ be an independent identically distributed sample from (X, Y), with the typical model of the form:

$$Y_i = m(X_i) + \sigma(X_i) e_i, \qquad \mathbb{E}[e_i] = 0, \quad \text{Var}(e_i) = 1$$

and X and e are independent. Also, for my own sanity, note: $\mathbb{X} = (X_1, \cdots, X_n)$.

The order-p local polynomial at x_0 minimizes the objective function of the form (and assuming that the (p+1)th derivative of the regression function m at point x_0 exists):

$$\sum_{i=1}^{n} \left(Y_i - \sum_{j=0}^{p} \beta_j (X_i - x_0)^j \right)^2 K_h(X_i - x_0) \qquad K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right)$$

Let X be the $n \times (p+1)$ local design matrix with $(j+1)^{\text{st}}$ column $(X_i - x_0)^j$,

let $W = \text{diag}\{K_h(X_i - x_0)\}$, and define

$$S_n = X^\top W X, \qquad \hat{\beta} = (X^\top W X)^{-1} X^\top W Y = S_n^{-1} X^\top W Y$$

We then turn back to our objective function (an estimator of $m^{(\nu)}(x_0)$):

$$\hat{m}^{(\nu)}(x_0) = \sum_{i=1}^{n} \left(Y_i - \sum_{j=0}^{p} \beta_j (X_i - x_0)^j \right)^2 K_h(X_i - x_0)$$

where $\varepsilon_{\nu+1}$ is the $(\nu+1)$st canonical basis vector.

A $(p+1)$-term Taylor expansion of m at x_0 gives

$$m(X_i) = \sum_{j=0}^{p} \beta_j (X_i - x_0)^j + r_i, \qquad \beta_j = \frac{m^{(j)}(x_0)}{j!},$$

with remainder

$$r_i = \frac{m^{(p+1)}(x_0)}{(p+1)!} (X_i - x_0)^{p+1} + o\big(|X_i - x_0|^{p+1}\big)$$

Then

$$\mathbb{E}[\hat{\beta} \mid \mathbb{X}] = (X^\top W X)^{-1} X^\top W m = \beta + S_n^{-1} X^\top W r \Rightarrow \text{bias}[\hat{\beta} \mid \mathbb{X}] = S_n^{-1} X^\top W r$$

Let $S_{n,j} = \sum_{i=1}^n (X_i - x_0)^j K_h(X_i - x_0)$ and $c_n = (S_{n,p+1}, \ldots, S_{n,2p+1})^\top$.
Using the remainder,

$$X^\top W r = \beta_{p+1}\, c_n + o_p\big((nh^{p+1}, \ldots, nh^{2p+1})^\top\big), \qquad \beta_{p+1} = \frac{m^{(p+1)}(x_0)}{(p+1)!}$$

Under standard regularity (f_X, σ continuous at x_0, $h \to 0$, $nh \to \infty$), define

$$H = \text{diag}(1, h, \ldots, h^p), \qquad \mu_j = \int u^j K(u)\, du$$

$$\mathbf{S} = (\mu_{j+\ell})_{0 \le j,\ell \le p}, \qquad \mathbf{c}_p = (\mu_{p+1}, \ldots, \mu_{2p+1})^\top,$$

and approximate

$$S_n \approx n f_X(x_0)\, H \mathbf{S} H, \qquad c_n \approx n f_X(x_0)\, H \mathbf{c}_p\, h^{p+1}$$

Substitute into the prior equation to get:

$$\text{bias}[\hat{\beta} \mid \mathbb{X}] = H^{-1} \mathbf{S}^{-1} \mathbf{c}_p\, \beta_{p+1}\, h^{p+1}\, [1 + o_p(1)]$$

Project to the νth derivative:

$$\begin{aligned} \text{bias}[\hat{m}_\nu(x_0) \mid \mathbb{X}] &= \nu!\, \varepsilon_{\nu+1}^\top H^{-1} \mathbf{S}^{-1} \mathbf{c}_p\, \beta_{p+1}\, h^{p+1} \\ &= \varepsilon_{\nu+1}^\top \mathbf{S}^{-1} \mathbf{c}_p\, \frac{\nu!}{(p+1)!}\, m^{(p+1)}(x_0)\, h^{p+1-\nu}\, [1 + o_p(1)] \end{aligned}$$

For $p - \nu$ odd, the leading term $\mathbf{c}_p$ does not cancel, so we're left with:

$$\text{bias}[\hat{m}_\nu(x_0) \mid \mathbb{X}] = \varepsilon_{\nu+1}^\top \mathbf{S}^{-1} \mathbf{c}_p\, \frac{\nu!}{(p+1)!}\, m^{(p+1)}(x_0)\, h^{p+1-\nu} + o_p\left(\frac{1}{nh^{1+2\nu}}\right)$$

Define the equivalent kernel functions:

$$K^\star_{\nu,p}(t) = \varepsilon_{\nu+1}^\top \mathbf{S}^{-1}\, (1, t, \cdots, t^p)^\top\, K(t)$$

Then (random design with density f_X),

$$\hat{m}_\nu(x_0) = \frac{1}{nh^{\nu+1} f_X(x_0)} \sum_{i=1}^n K^\star_{\nu,p}\left(\frac{X_i - x_0}{h}\right) Y_i\, [1 + o_p(1)]$$

This kernel satisfies the moment conditions:

$$\int u^q K^\star_{\nu,p}(u), du = \delta_{\nu q} \qquad 0 \le \nu, q \le p$$

Therefore,

$$\operatorname{bias}[\hat{m}_\nu(x_0) \mid \mathbb{X}] = \left(\int t^{p+1} K^\star_{\nu,p}(t)\, dt \right) \frac{\nu!}{(p+1)!}\, m^{(p+1)}(x_0)\, h^{p+1-\nu} + o_p\big(h^{p+1-\nu}\big)$$

For $p - \nu$ odd.

2.

Write the conditional variance of the local polynomial regression estimator

$$\mathrm{Var}[\hat{m}_\nu(x_0) \mid \mathbb{X}] = \varepsilon_{\nu+1}^T \mathbf{S}^{-1}\mathbf{S}^\star\mathbf{S}^{-1}\varepsilon_{\nu+1}\frac{\nu!^2\sigma^2(x_0)}{f_X(x_0)nh^{1+2\nu}} + o_p\left(\frac{1}{nh^{1+2\nu}}\right)$$

in terms of the equivalent kernel $K^\star_{\nu,p}$ (see p. 60 Eq. (4.30) in the notes).

Answer

Under the same setup and assumptions of Question 1, we have:

$$\hat{m}_\nu(x_0) = \nu!\,\varepsilon_{\nu+1}^\top\hat{\beta}$$

Conditional on X, the error covariance is
$\mathrm{Var}(Y \mid \mathbb{X}) = \mathrm{diag}\{\sigma^2(X_i)\}$, so

$$\mathrm{Var}[\hat{m}_\nu(x_0) \mid \mathbb{X}] = \nu!^2\,\varepsilon_{\nu+1}^\top S_n^{-1}X^\top W\,\mathrm{Var}(Y \mid \mathbb{X})WXS_n^{-1}\varepsilon_{\nu+1}$$

Assuming $\sigma^2(X_i) \approx \sigma^2(x_0)$ near x_0,

$$\mathrm{Var}[\hat{m}_\nu(x_0) \mid \mathbb{X}] = \nu!^2\sigma^2(x_0)\varepsilon_{\nu+1}^\top S_n^{-1}(X^\top W^2X)S_n^{-1}\varepsilon_{\nu+1}\,[1 + o_p(1)]$$

Let

$$H = \mathrm{diag}(1, h, \dots, h^p), \quad \mu_j = \int u^j K(u)\,du, \quad \mu_j^{(2)} = \int u^j K(u)^2\,du,$$

and define

$$\mathbf{S} = (\mu_{j+\ell})_{0\le j,\ell\le p} \qquad \mathbf{S}^{(2)} = (\mu_{j+\ell}^{(2)})_{0\le j,\ell\le p}$$

With f_X continuous at x_0,

$$S_n \approx nf_X(x_0)H\mathbf{S}H, \qquad X^\top W^2X \approx \frac{nf_X(x_0)}{h}H\mathbf{S}^{(2)}H$$

Hence

$$S_n^{-1} \approx \frac{1}{nf_X(x_0)}H^{-1}\mathbf{S}^{-1}H^{-1}$$

Substitute into the previous equation to get:

$$\mathrm{Var}[\hat{m}_\nu(x_0) \mid \mathbb{X}] = \nu!^2\sigma^2(x_0)\frac{1}{nf_X(x_0)}\frac{1}{h}\,\varepsilon_{\nu+1}^\top\left[H^{-1}\mathbf{S}^{-1}\mathbf{S}^{(2)}\mathbf{S}^{-1}H^{-1}\right]\varepsilon_{\nu+1}\,[1 + o_p(1)]$$

Since $H^{-1}\varepsilon_{\nu+1} = h^{-\nu}\varepsilon_{\nu+1}$,

$$\mathrm{Var}[\hat{m}_\nu(x_0) \mid \mathbb{X}] = \varepsilon_{\nu+1}^\top \mathbf{S}^{-1}\mathbf{S}^{(2)}\mathbf{S}^{-1}\varepsilon_{\nu+1} \frac{\nu!^2\sigma^2(x_0)}{f_X(x_0)\, n\, h^{1+2\nu}} + o_p\left(\frac{1}{nh^{1+2\nu}}\right)$$

Define the equivalent kernel functions as before:

$$K^\star_{\nu,p}(t) = \varepsilon_{\nu+1}^\top \mathbf{S}^{-1}\left(1, t, \cdots, t^p\right)^\top K(t)$$

Then, under random design with density f_X,

$$\hat{m}_\nu(x_0) = \frac{1}{nh^{\nu+1} f_X(x_0)} \sum_{i=1}^{n} K^\star_{\nu,p}\left(\frac{X_i - x_0}{h}\right) Y_i[1 + o_p(1)]$$

Using $\mathrm{Var}(Y_i \mid \mathbb{X}) \approx \sigma^2(x_0)$ and the independence of Y_i,

$$\mathrm{Var}[\hat{m}_\nu(x_0) \mid \mathbb{X}] = \frac{\sigma^2(x_0)}{nh^{1+2\nu} f_X(x_0)} \int [K^\star_{\nu,p}(t)]^2\, dt + o_p\left(\frac{1}{nh^{1+2\nu}}\right)$$

From the above, combining terms,

$$\varepsilon_{\nu+1}^\top \mathbf{S}^{-1}\mathbf{S}^\star\mathbf{S}^{-1}\varepsilon_{\nu+1} = \int [K^\star_{\nu,p}(t)]^2\, dt, \qquad \mathbf{S}^\star = \mathbf{S}^{(2)}$$

Note: This result also relies upon WLS representation and variance in addition to the results noted in the book.

3.

Show that the equivalent kernel satisfies the following moment condition

$$\int u^q K^{\star}_{\nu,p}(u)\, du = \delta_{\nu,q} \quad 0 \leq \nu, q \leq p,$$

where $\delta_{\nu,q} = 1$ if $\nu = q$ and 0 else.

Answer

Let

$$v_p(u) = (1, u, \dots, u^p)^\top, \qquad \mathbf{S} = \int v_p(u)\, v_p(u)^\top K(u)\, du = \big(\mu_{j+\ell}\big)_{0 \leq j, \ell \leq p}$$

where $\mu_r = \int u^r K(u)\, du$.

Recall the equivalent kernel

$$K^{\star}_{\nu,p}(u) = \varepsilon_{\nu+1}^\top \mathbf{S}^{-1} v_p(u)\, K(u)$$

with $\varepsilon_{\nu+1}$ the $(\nu+1)$-st canonical basis vector in $\mathbb{R}^{p+1}$.

For $0 \leq q \leq p$, compute the q-th moment of $K^{\star}_{\nu,p}$:

$$\int u^q K^{\star}_{\nu,p}(u)\, du = \varepsilon_{\nu+1}^\top \mathbf{S}^{-1} \left(\int v_p(u)\, u^q K(u)\, du \right)$$

Define the vector

$$s_q = \int v_p(u)\, u^q K(u)\, du = \big(\mu_q \;\; \mu_{q+1} \;\vdots\; \mu_{q+p}\big)$$

Observe that s_q is exactly the $(q+1)$-st column of $\mathbf{S}$:

for $j = 0, \dots, p$,

$$(s_q)_{j+1} = \mu_{q+j} = \mathbf{S}_{j+1,q+1}$$

Hence $s_q = \mathbf{S}, \varepsilon_{q+1}$. Therefore,

$$\int u^q K^{\star}_{\nu,p}(u)\, du = \varepsilon_{\nu+1}^\top \mathbf{S}^{-1}\, \mathbf{S}\, \varepsilon_{q+1} = \varepsilon_{\nu+1}^\top \varepsilon_{q+1} = \delta_{\nu,q}$$

which proves the stated moment condition for all $0 \leq \nu, q \leq p$.

4.

Show that the weights W_ν^n satisfy the following discrete moment condition

$$\sum_{i=1}^{n} (X_i - x_0)^q W_\nu^n \left(\frac{X_i - x_0}{h} \right) = \delta_{\nu,q} \quad 0 \le \nu, q \le p$$

Answer

Let

$$x_i = \begin{pmatrix} 1 & (X_i - x_0) & \cdots & (X_i - x_0)^p \end{pmatrix}^\top, \qquad X = \begin{pmatrix} x_1^\top \vdots x_n^\top \end{pmatrix}, \qquad W = \operatorname{diag}\big(K_h(X_i - x_0)\big)$$

Define

$$S_n = X^\top W X$$

Then, the order-p local polynomial estimator of the νth derivative at x_0 can be written in **linear smoother form** as

$$\hat{m} * \nu(x_0) = \sum_{i=1}^{n} W_\nu^n \left(\frac{X_i - x_0}{h} \right) Y_i$$

where the **weights** are defined (for $0 \le \nu \le p$) as

$$W_\nu^n \left(\frac{X_i - x_0}{h} \right) = \varepsilon_{\nu+1}^\top S_n^{-1} X^\top W e_i = \varepsilon_{\nu+1}^\top S_n^{-1} x_i \, K_h(X_i - x_0)$$

and $\varepsilon_{\nu+1}$ is the $(\nu+1)$st canonical basis vector, while e_i is the ith standard basis vector in $\mathbb{R}^n$.

Fix $q \in 0, 1, \ldots, p$. We wish to show that the local polynomial weights satisfy

$$\sum_{i=1}^{n} (X_i - x_0)^q \, W_\nu^n \left(\frac{X_i - x_0}{h} \right) = \delta_{\nu,q}$$

Substituting the definition of W_ν^n,

$$\begin{aligned} \sum_{i=1}^{n} (X_i - x_0)^q \, W_\nu^n \left(\frac{X_i - x_0}{h} \right) &= \sum_{i=1}^{n} (X_i - x_0)^q \, \varepsilon_{\nu+1}^\top S_n^{-1} x_i \, K_h(X_i - x_0) \\ &= \varepsilon_{\nu+1}^\top S_n^{-1} \left(\sum_{i=1}^{n} x_i (X_i - x_0)^q K_h(X_i - x_0) \right) \end{aligned}$$

Note that

$$\sum_{i=1}^{n} x_i (X_i - x_0)^q K_h(X_i - x_0) = \sum_{i=1}^{n} x_i x_i^\top K_h(X_i - x_0) \, \varepsilon_{q+1} = S_n \, \varepsilon_{q+1}$$

Substituting this identity back gives

$$\begin{aligned}\sum_{i=1}^{n}(X_i - x_0)^q \, W_\nu^n\left(\frac{X_i - x_0}{h}\right) &= \varepsilon_{\nu+1}^\top S_n^{-1}(S_n \varepsilon_{q+1}) \\ &= \varepsilon_{\nu+1}^\top \varepsilon_{q+1} \\ &= \delta_{\nu,q}\end{aligned}$$

$$\sum_{i=1}^{n}(X_i - x_0)^q \, W_\nu^n\left(\frac{X_i - x_0}{h}\right) = \delta_{\nu,q} \qquad 0 \leq \nu, q \leq p$$

This **discrete moment condition** shows that the local polynomial regression weights exactly reproduce monomials up to degree p. That is, the weights annihilate all lower-order polynomial components except for the one corresponding to the νth derivative. Consequently, the local polynomial estimator $\hat{m}_\nu(x_0)$ isolates the νth derivative of $m(x)$ at x_0, ensuring unbiasedness for all polynomials of degree $\leq p$.