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 i.i.d. sample from (X, Y), with the typical model of the form:

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

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

The order-p local polynomial estimator at x_0 minimizes:

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

Let $\mathbf{X}$ be the $n \times (p+1)$ design matrix whose $(j+1)$-st (-th?) column has entries $(X_i - x_0)^j$, $i = 1, \cdots, n$ and $j = 1, \cdots p$.

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

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

The estimator of the νth derivative is:

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

A $(p+1)$-term Taylor expansion 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 term r_i given by:

$$r_i = \frac{m^{(p+1)}(x_0)}{(p+1)!} (X_i - x_0)^{p+1} + o_p(h^{p+1})$$

Thus:

$$\mathbb{E}[\hat{\beta} \mid \mathbb{X}] = \beta + S_n^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{r}, \qquad \text{bias}[\hat{\beta} \mid \mathbb{X}] = S_n^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{r}$$

Let:

$$S_{n,j} = \sum_{i=1}^{n} (X_i - x_0)^j K_h(X_i - x_0), \qquad c_n = (S_{n,p+1}, \dots, S_{n,2p+1})^\top$$

Using the remainder term:

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

Define:

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

And additionally:

$$S = (\mu_{j+\ell})_{0 \le j, \ell \le p} \qquad \text{and } c_p = (\mu_{p+1}, \dots, \mu_{2p+1})^\top$$

Note: There are two different moment vectors in this derivation: $c_n = (S_{n,p+1}, \dots, S_{n,2p+1})^\top$ and $c_p = (\mu_{p+1}, \dots, \mu_{2p+1})^\top$.

The text also uses both quantities, but only explicitly defines c_n. Further, c_n denotes the sample vector of weighted empirical moments while c_p is the population moment vector appearing in the asymptotic bias formula.

Continuing on, for continuous f_X at x_0,

$$S_n \approx n f_X(x_0)\, HSH, \qquad c_n \approx n f_X(x_0)\, H c_p\, h^{p+1} \tag{4.19}$$

Taken together:

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

Projecting to the νth derivative:

$$\mathrm{bias}[\hat{m}_\nu(x_0) \mid \mathbb{X}] = \nu!\, \varepsilon_{\nu+1}^\top H^{-1} S^{-1} c_p\, \beta_{p+1}\, h^{p+1}$$

so

$$\mathrm{bias}[\hat{m}_\nu(x_0) \mid \mathbb{X}] = \varepsilon_{\nu+1}^\top S^{-1} c_p \frac{\nu!}{(p+1)!} m^{(p+1)}(x_0)\, h^{p+1-\nu} + o_p(h^{p+1-\nu})$$

For $p - \nu$ odd, the leading term does not vanish.

Now define the equivalent kernel:

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

As given, the equivalent kernel satisfies the moment conditions:

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

Also:

$$\int t^{p+1} K^{\star}_{\nu,p}(t)\, dt = \varepsilon^{\top}_{\nu+1} S^{-1} c_p$$

Substituting this identity into the bias expression gives:

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

for $p - \nu$ odd (matching the equation given in 4.29).

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 $\mathbb{X}$:

$$\mathrm{Var}(Y \mid \mathbb{X}) = \mathrm{diag}\{\sigma^2(X_i)\}$$

(which assumes independence of errors)

So:

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

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

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

Let:

$$H = \mathrm{diag}(1, h, \ldots, h^p), \qquad \mu_j = \int u^j K(u)\,du, \qquad \text{and } \nu_j = \mu_j^{(2)} = \int u^j K(u)^2\,du$$

And define:

$$\mathbf{S} = (\mu_{j+\ell})_{0\le j,\ell\le p}, \qquad \mathbf{S}^\star = (\mu^{(2)}_{j+\ell})_{0\le j,\ell\le p} = \begin{pmatrix} \nu_0 & \nu_1 & \cdots & \nu_p \\ \nu_1 & \nu_2 & \cdots & \nu_{p+1} \\ \vdots & \vdots & \ddots & \vdots \\ \nu_p & \nu_{p+1} & \cdots & \nu_{2p} \end{pmatrix} \tag{4.20}$$

Note: For what follows, I use ν_j in lieu of $\mu^{(2)}$ for notational convenience.

With f_X continuous at x_0:

$$S_n \approx nf_X(x_0)\,H\mathbf{S}H, \qquad \mathbf{X}^\top\mathbf{W}^2\mathbf{X} \approx \frac{nf_X(x_0)}{h}\,H\mathbf{S}^\star H \qquad \text{(Also 4.20)}$$

Hence

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

Substituting,

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

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

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

Now define the equivalent kernel:

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

Under random design with density f_X:

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

Using the following:

$$\text{Var}(Y_i \mid \mathbb{X}) \approx \sigma^2(x_0),$$

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

And, under the assumption of independence (specifically independence of errors conditional on $\mathbb{X}$), we then have:

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

Matching the equation given in 4.30.

3.

Show that the equivalent kernel satisfies the following moment condition

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

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

Answer

Let:

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

Where we follow the typical convention: $\mu_r = \int u^r K(u)\,du$.

As defined, the equivalent kernel is of the form:

$$K^\star_{\nu,p}(u) = \varepsilon_{\nu+1}^\top \mathbf{S}^{-1} v_p(u)\, K(u) \qquad \text{(As in Eq. 4.27, but using u instead of t)}$$

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

For $0 \le q \le p$, the q-th moment of $K^\star_{\nu,p}$ is:

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

Define the vector

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

Note: s_q is exactly the $(q+1)$-st column of $\mathbf{S}$, i.e., 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:

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

Noting that:

$$\varepsilon_{\nu+1}^\top \varepsilon_{q+1} = \begin{cases} 1, & \nu = q, \\ 0, & \nu \ne q, \end{cases}$$

We have then proved that the equivalent kernel satisfies the moment condition

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

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

Note on notation: $\mathbf{X}$ and $\mathbf{W}$ denote matrices, and $W_\nu^n(\cdot)$ denotes the scalar weight function used in the local polynomial estimator (so that $W_\nu^n((X_i - x_0)/h)$ is the weight applied to observation i). I attempted to follow the notation of the text throughout.

That all being said, let:

$$\mathbf{X} = \begin{pmatrix} 1 & (X_1 - x_0) & \cdots & (X_1 - x_0)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (X_n - x_0) & \cdots & (X_n - x_0)^p \end{pmatrix}, \qquad \mathbf{W} = \operatorname{diag}\big(K_h(X_i - x_0)\big)$$

And define:

$$S_n = \mathbf{X}^\top \mathbf{W} \mathbf{X}$$

Then, the order-p local polynomial estimator of the νth derivative at x_0 can be written as:

$$\hat{m}_\nu(x_0) = \sum_{i=1}^{n} W_\nu^n\left(\frac{X_i - x_0}{h}\right) Y_i \qquad \text{(4.25 of the text)}$$

Where the weights are defined (for $0 \le \nu \le p$) as:

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

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 $q \in \{0, 1, \ldots, p\}$.

The goal is to show:

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

To that end, substituting the definition of W_ν^n given above:

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

Note that

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

Where, notationally (equivalent to the setup):

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

And where:

$$x_i^\top \varepsilon_{q+1} = (X_i - x_0)^q$$

Substituting this back into the overall equation gives:

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

Thus, the weights W_ν^n satisfy the discrete moment condition:

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