STAT 526 Smoothing Methods in Regression — Term 2, 1998/99
February 23, 1999

Local Polynomial Least Squares Estimate

1 Definition of the Estimate

Suppose, as usual, that the data are x_i, Y_i, $i = 1, \cdots, n$, with $Y_i = m(x_i) + \epsilon_i$, ϵ_i's independent with mean 0 and variance σ^2, $x_i \in [0,1]$. Assume that the x_i's follow a regular design with positive design density f (see the Nadaraya-Watson handout for the definition of a regular design).

Choose p, the degree of the polynomial to be fit. Fix x and let $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_x = (\hat{\beta}_0 \cdots, \hat{\beta}_p)^t$ be the minimizer of

$$\sum_{i=1}^{n} [Y_i - \beta_0 - \beta_1(x_i - x) - \cdots - \beta_p(x_i - x)^p]^2 \ K\left(\frac{x_i - x}{h}\right). \qquad (1)$$

We then take

$$\hat{m}^{(k)}(x) = k!\hat{\beta}_k, \quad k = 0, \cdots, p.$$

When $k = 0$, $\hat{m}^{(k)}(x) = \hat{m}(x) = \hat{\beta}_0$.

Section 2 contains a general expression for the asymptotic bias and variance of $\hat{m}^{(k)}(x)$. In Section 3, these expressions are written in terms of a new kernel, the *effective kernel*, $K_{j,p}$.

2 Asymptotic Bias and Variance

Theorem 1 below gives a very general form of the asymptotic bias and of the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}_x$. The Corollary looks at the specific case of $p = 1$, fitting a local linear estimate. Note that the case $p = 0$ yields the Nadaraya-Watson estimate of $m(x)$.

The following conditions and definitions are needed for the asymptotic results. Suppose that

1. K is Lipshitz and has support $[-1, 1]$;

2. f' is Lipshitz on $[0,1]$;

3. m has $p+2$ continuous derivatives;

4. Let $X = X_x$ be the n by $p+1$ matrix with $X_{ij} = (x_i - x)^{j-1}$. Let $K_h = K_{hx}$ be n by n diagonal with $[K_h]_{ii} = K((x_i - x)/h)$. Suppose that $X^t K_h X$ is a positive definite matrix.

Using matrix and vector notation, we see that minimizing (1) is equivalent to minimizing

$$(\mathbf{Y} - X\boldsymbol{\beta})^t K_h (\mathbf{Y} - X\boldsymbol{\beta})$$

over $\boldsymbol{\beta}$, where $\mathbf{Y} = (Y_1, \cdots, Y_n)^t$. Differentiating with respect to $\boldsymbol{\beta}$ and setting equal to 0 yields the $p+1$ equations:

$$-2X^t K_h (\mathbf{Y} - X\hat{\boldsymbol{\beta}}) = 0$$

or

$$\hat{\boldsymbol{\beta}} = [X^t K_h X]^{-1} X^t K_h \mathbf{Y}.$$

By taking second derivatives, we see that $\hat{\boldsymbol{\beta}}$ is the unique minimizer since $X^t K_h X$ is a positive definite matrix.

Recall that

$$\nu_L(K, x, h) = \int_{s \in [-x/h, (1-x)/h]} s^L \ K(s) \ ds$$

and that, if $x \in [h, 1-h]$ and K has support $[-1, 1]$, then $\nu_L(K, x, h) = \int_{-1}^{1} s^L \ K(s) \ ds \equiv \nu_L(K)$. Let

$$g_1(x) = \frac{m^{(p+1)}(x)}{(p+1)!}$$

and

$$g_2(x) = \frac{f'(x)}{f(x)} \frac{m^{(p+1)}(x)}{(p+1)!} + \frac{m^{(p+2)}(x)}{(p+2)!}.$$

Let A, B, and C $p+1$ by $p+1$ matrices with

$$A_{ij} = A_{ij}(x, h) = \nu_{i+j-2}(K, x, h),$$

$$B_{ij} = B_{ij}(x, h) = \nu_{i+j-1}(K, x, h),$$

$$C_{ij} = C_{ij}(x,h) = \nu_{i+j-2}(K^2, x, h),$$

and $\mathbf{v}$ and $\mathbf{w}$ $p+1$ vectors with ith components

$$\mathbf{v}_i = \nu_{p+i}(K, x, h), \qquad \mathbf{w}_i = \nu_{p+i+1}(K, x, h).$$

Theorem 1 *Assume the conditions 1-4 above hold. Let $\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}_x$ be a $p+1$ vector with ith component equal to $h^{i-1}[\hat{\beta}_{i-1} - m^{(i-1)}(x)/(i-1)!]$. If $h \to 0$ and $nh^3 \to \infty$, then*

$$\mathrm{E}[\tilde{\boldsymbol{\beta}}_x] = h^{p+1} g_1(x) A^{-1}\mathbf{v}$$
$$+ h^{p+2}\left[g_2(x)A^{-1}\mathbf{w} - \frac{f'(x)}{f(x)} g_1(x) A^{-1} B A^{-1}\mathbf{v}\right] + o(h^{p+2})$$

and

$$\mathrm{cov}[\tilde{\boldsymbol{\beta}}_x] = \frac{\sigma^2}{nhf(x)} A^{-1} C A^{-1} + o\left(\frac{1}{nh}\right)$$

uniformly in $x \in [0,1]$.

The Theorem is proved in Section 5. Applying the Theorem to specific cases is straightforward but tedious. In the following Corollary, the Theorem is used to derive the results for the local linear estimate.

Corollary 1 *Suppose that $p = 1$ and that K is a probability density function symmetric about 0. In addition, assume conditions 1-4 and that $h \to 0$ and $nh^3 \to \infty$. Then, uniformly for $x \in [h, 1-h]$*

$$\mathrm{E}[\hat{m}(x)] - m(x) = h^2\ \frac{m''(x)}{2}\nu_2(K) + O(h^3),$$

$$\mathrm{E}[\hat{m}'(x)] - m'(x) = h^2\left[\frac{f'(x)}{f(x)}\frac{m''(x)}{2} + \frac{m'''(x)}{6}\right]\frac{\nu_4(K)}{\nu_2(K)}$$
$$-h^2\frac{f'(x)}{f(x)}\frac{m''(x)}{2}\nu_2(K) + o(h^2),$$

$$\mathrm{var}[\hat{m}(x)] = \frac{\sigma^2}{nhf(x)}\nu_0(K^2) + o\left(\frac{1}{nh}\right)$$

and

$$\mathrm{var}[\hat{m}'(x)] = \frac{\sigma^2}{nh^3 f(x)} \frac{\nu_2(K^2)}{\nu_2^2(K)} + o\left(\frac{1}{nh^3}\right).$$

Uniformly in $x \notin [h, 1-h]$

$$\mathrm{E}[\hat{m}(x) - m(x)] = h^2 \frac{m''(x)}{2} [A^{-1}\mathbf{v}]_1 + O(h^3)$$

$$\mathrm{E}[\hat{m}'(x) - m'(x)] = h \frac{m''(x)}{2} [A^{-1}\mathbf{v}]_2 + O(h^2)$$

$$var[\hat{m}(x)] = \frac{[A^{-1}CA^{-1}]_{11}}{nh} + o\left(\frac{1}{nh}\right),$$

$$\mathrm{var}[\hat{m}'(x)] = \frac{[A^{-1}CA^{-1}]_{22}}{nh^3} + o\left(\frac{1}{nh^3}\right)$$

Comments on the Corollary. Consider the bias in estimating $m(x)$. For $x \in [h, 1-h]$, when $m''(x) \neq 0$, the bias is of order h^2, just as with the Nadaraya-Watson (local constant) estimate. For $x \notin [h, 1-h]$, we see that the bias of the local linear estimate is still of order h^2. Recall that the Nadaraya-Watson estimate has bias of order h at the boundary. Both estimates of $m(x)$ have variance of order $1/(nh)$ for $x \in [0, 1]$.

So with the local linear estimate of m we have no boundary problem with the MSE, that is

$$IMSE(h) = \int_0^1 \mathrm{E}[\hat{m}(x) - m(x)]^2\, dx \sim h^4 \int_0^1 B^2(x)\, dx + \frac{\sigma^2}{nh} \int_0^1 V(x)\, dx,$$

where

$$B(x) = m''(x)\nu_2(K)/2$$

and

$$V(x) = \frac{\sigma^2}{f(x)} \nu_0(K^2).$$

Now consider the estimate of $m'(x)$. For $x \in [h, 1-h]$ the bias is of order h^2. However, for $x \notin [h, 1-h]$, the bias is of order h. One can show that the $IMSE$ of $\hat{m}'$ (that is $\int_0^1 \mathrm{E}[\hat{m}'(x) - m'(x)]^2\, dx$) is dominated by the square of the bias of $\hat{m}'(x)$ for $x \notin [h, 1-h]$, that is by $\int_{x \notin [h,1-h]} \mathrm{E}[\hat{m}'(x) - m'(x)]^2 \, dx$. This parallels the properties of the Nadaraya-Watson estimate of m where the $IMSE$ of $\hat{m}$ is dominated by the square of the bias of $\hat{m}(x)$ for $x \notin [h, 1-h]$.

Note that the asymptotic variance of $\hat{m}'(x)$ is of order $1/(nh^3)$, which is the same order as for the Nadaraya-Watson based estimates of $m'(x)$. The order $1/(nh^3)$ is larger than $1/(nh)$, the order of the asymptotic variance of $\hat{m}(x)$. Typically a derivative is harder to estimate than the function itself, and the variance expressions quantify the problem.

Proof of the Corollary. By Theorem 1

$$\mathrm{E}\begin{pmatrix} \hat{m}(x) - m(x) \\ h[\hat{m}'(x) - m'(x)] \end{pmatrix} = h^2 \, \frac{m''(x)}{2} A^{-1}\mathbf{v}$$

$$+h^3 \left[\frac{f'(x)}{f(x)} \frac{m''(x)}{2} + \frac{m'''(x)}{6} \right] A^{-1}\mathbf{w}$$

$$-h^3 \frac{f'(x)}{f(x)} \frac{m''(x)}{2} A^{-1} B A^{-1}\mathbf{v} + o(h^3).$$

Since K is a probability density symmetric about 0, $\nu_0(K) = 1$ and $\nu_j(K) = 0 = \nu_j(K^2)$, j odd. So, for $x \in [h, 1-h]$,

$$A = \begin{bmatrix} 1 & 0 \\ 0 & \nu_2(K) \end{bmatrix} \qquad B = \begin{bmatrix} 0 & \nu_2(K) \\ \nu_2(K) & 0 \end{bmatrix} \qquad C = \begin{bmatrix} \nu_0(K^2) & 0 \\ 0 & \nu_2(K^2) \end{bmatrix}$$

$$\mathbf{v} = \begin{pmatrix} \nu_2(K) \\ 0 \end{pmatrix} \qquad \mathbf{w} = \begin{pmatrix} 0 \\ \nu_4(K) \end{pmatrix}$$

and

$$\mathrm{E}\begin{pmatrix} \hat{m}(x) - m(x) \\ h[\hat{m}'(x) - m'(x)] \end{pmatrix} = h^2 \, \frac{m''(x)}{2} \begin{pmatrix} \nu_2(K) \\ 0 \end{pmatrix}$$

$$+h^3 \left[\frac{f'(x)}{f(x)} \frac{m''(x)}{2} + \frac{m'''(x)}{6} \right] \begin{pmatrix} 0 \\ \nu_4(K)/\nu_2(K) \end{pmatrix}$$

$$-h^3 \frac{f'(x)}{f(x)} \frac{m''(x)}{2} \begin{pmatrix} 0 \\ \nu_2(K) \end{pmatrix} + o(h^3).$$

So

$$\mathrm{E}[\hat{m}(x) - m(x)] = h^2 \, \frac{m''(x)}{2} \nu_2(K) + O(h^3),$$

$$h\mathrm{E}[\hat{m}'(x) - m'(x)] = h^3 \left[\frac{f'(x)}{f(x)} \frac{m''(x)}{2} + \frac{m'''(x)}{6} \right] \frac{\nu_4(K)}{\nu_2(K)}$$

$$-h^3 \frac{f'(x)}{f(x)} \frac{m''(x)}{2} \nu_2(K) + o(h^3).$$

Now consider the covariance matrix of $\tilde{\boldsymbol{\beta}}$ for $x \in [h, 1-h]$.

$$\begin{aligned} \mathrm{cov}(\tilde{\boldsymbol{\beta}}) &= \frac{\sigma^2}{nhf(x)} A^{-1} C A^{-1} + o\left(\frac{1}{nh}\right) \\ &= \frac{\sigma^2}{nhf(x)} \begin{bmatrix} \nu_0(K^2) & 0 \\ 0 & \nu_2(K^2)/\nu_2^2(K) \end{bmatrix} + o\left(\frac{1}{nh}\right). \end{aligned}$$

Therefore

$$\mathrm{var}[\hat{m}(x)] = \frac{\sigma^2}{nhf(x)} \nu_0(K^2) + o\left(\frac{1}{nh}\right)$$

and

$$\mathrm{var}[h\hat{m}'(x)] = \frac{\sigma^2}{nhf(x)} \frac{\nu_2(K^2)}{\nu_2^2(K)} + o\left(\frac{1}{nh}\right).$$

The calculations for $x \notin [h, 1-h]$ are similar.

References. The asymptotic means and variances for the local linear least squares estimates were derived by Müller (1987, *JASA*) for the nonrandom design case. Ruppert and Wand (1994, *Annals of Statistics*) considered the case of a random design ($X_1, \cdots, X_n$ iid), general degree polynomial, x's possible vectors. The proof of the Theorem (given in Section 5) is a modification of the proof of Ruppert's and Wand's result, but for the fixed design case, $x_i \in \Re$. The random design and the fixed design have the same asymptotic bias and variance.

Exercise 1. Suppose that $K(u) = (2\pi)^{-1/2} \exp(-u^2/2)$. Using Corollary 1 find the asymptotic bias and variance of the local linear estimate of $m(x)$.

3 The Effective Kernel $K_{j,p}$

The asymptotic bias and variance of the local polynomial estimate depend on K in a rather complicated way. However, this dependence can be re-written in terms of another kernel $K_{j,p}$ defined below. Writing things in terms of the kernel $K_{j,p}$ simplifies notation and also provides some insight into the properties of the estimate. This work is from Ruppert and Wand (1994, *Annals of Statsitics*).

Recall the definition of A from Section 2: $A_{ij} = A_{ij}(x,h) = \nu_{i+j-2}(K,x,h)$. Fix j and p and let $M_{j,p}$ be the $p+1$ by $p+1$ matrix which is equal to A except that the $(j+1)$st column is replaced by $(1, u, u^2, \cdots, u^p)^t$. Let

$$K_{j,p}(u) = j!\frac{\det\ M_{j,p}(u)}{\det\ A}K(u).$$

The following theorem gives the asymptotic bias and variance of $\hat{m}^{(j)}$ in terms of $K_{j,p}$.

Theorem 2 *Suppose that the conditions of Theorem 1 hold. Then*

$$\mathrm{var}(\hat{m}^{(j)}(x)) = \frac{\sigma^2}{nh^{2j+1}f(x)}\nu_0(K_{j,p}^2, x, h) + o\left(\frac{1}{nh^{2j+1}}\right)$$

uniformly in $x \in [0,1]$. Now suppose in addition that K is a probability density symmetric about zero. If $p-j$ is odd then, uniformly in $x \in [0,1]$

$$\mathrm{E}[\hat{m}^{(j)}(x)] - m^{(j)}(x) = h^{p-j+1}g_1(x)\nu_{p+1}\left(K_{j,p}, x, h\right) + o(h^{p-j+1}).$$

If $p-j$ is even then, uniformly in $x \in [h, 1-h]$,

$$\mathrm{E}[\hat{m}^{(j)}(x)] - m^{(j)}(x) = h^{p-j+2}\left[g_2(x)\nu_{p+2}\left(K_{j,p}\right) - \frac{f'(x)}{f(x)}g_1(x)j\nu_{p+1}\left(K_{j-1,p}\right)\right] + o(h^{p-j+2}).$$

If $p-j$ is even then uniformaly in $x \notin [h, 1-h]$

$$\mathrm{E}[\hat{m}^{(j)}(x)] - m^{(j)}(x) = h^{p-j+1}g_1(x)\nu_{p+1}\left(K_{j,p}, x, h\right) + o(h^{p-j+1}).$$

Exercise 2. Consider estimating $m'(x)$ by using a local quadratic. Find the asymptotic bias and variance of the estimate for $x \in [h, 1-h]$ in two ways: using Theorem 1 and then using Theorem 2. What is the bandwidth h_{opt} that minimizes the asymptotic integrated mean squared error? How might you choose a good value of h with from a data set? For this last question, no calculations are necessary, just describe what calculations you would need to make, what coding you would need to do. There may be several ways reasonable ways to choose h.

Comments on Theorem 2. Suppose K is a kernel that's a probability density symmetric about 0.

First consider estimating m with a local polynomial of degree p (so $j = 0$). If p is odd, the order of the bias of $\hat{m}(x)$ is h^{p+1} uniformly in $x \in [0, 1]$. If p is even, the order of the bias is h^{p+2} for x in the interior and h^{p+1} for x on the boundary. Since it's usually easier to use a global bandwidth, people often like to use p odd to estimate m using a global bandwidth. Typically p is chosen to be 1. Table 1 shows the orders of the asymptotic bias as a function of p.

TABLE 1. Order of Bias in $\hat{m}(x)$, a Local Polynomial Estimate

Degree p	$x \in [h, 1-h]$	$x \notin [h, 1-h]$
0	h^2	h
1	h^2	h^2
2	h^4	h^3
3	h^4	h^4
p odd	h^{p+1}	h^{p+1}
p even	h^{p+2}	h^{p+1}

Now consider estimating $m^{(j)}(x)$ with a local polynomial fit of degree $p \geq j$. If $p - j$ is odd, the order of the bias of $\hat{m}(x)$ is h^{p-j+1} uniformly in x. If $p - j$ is even, the order of the bias is h^{p-j+2} for x in the interior and h^{p-j+1} for x on the boundary. Thus the boundary bias dominates the integrated mean squared error. For this reason, people usually like to use $p - j$ odd to

TABLE 2. Order of Bias in $\hat{m}'(x)$, a Local Polynomial Estimate

Degree p	$x \in [h, 1-h]$	$x \notin [h, 1-h]$
1	h^2	h
2	h^2	h^2
3	h^4	h^3
4	h^4	h^4
p odd	h^{p+1}	h^p
p even	h^p	h^p

estimate $m^{(j)}$ using a global bandwidth. Typically $p = j + 1$. Table 2 shows the order of the bias for $\hat{m}'(x)$ $(j = 1)$.

The proof of Theorem 2 relies on the following Lemma, which is proven at the end of this section.

Lemma 1 *Recall the definitions of $A, \nu, \mathbf{v}$, and $\mathbf{w}$ from Section 2. Then*

1. $\nu_{p+1}(K_{i-1,p}, x, h) = (i-1)![A^{-1}\mathbf{v}]_i$ *and* $\nu_{p+2}(K_{i-1,p}, x, h) = (i-1)![A^{-1}\mathbf{w}]_i$*;*

2. *if K is a probability density symmetric about 0 and $x \in [h, 1-h]$ then $[A^{-1}\mathbf{v}]_i = 0$ for $i + p$ odd;*

3. *if K is a probability density symmetric about 0 and $x \in [h, 1-h]$ then $[A^{-1}\mathbf{w}]_i = 0$ for $i + p$ even;*

4. *if K is a probability density symmetric about 0 and $x \in [h, 1-h]$ then $[A^{-1}BA^{-1}\mathbf{v}]_i = 0$ for $i + p$ even, $= \nu_{p+1}(K_{i-2,p})/(i-2)!$ for $i + p$ odd.*

Exercise 3. Show that $\int_{u \in [-x/h, (1-x)/h]} K_{j,p}^2(u)\, du = (j!)^2 [A^{-1}CA^{-1}]_{j+1,j+1}$ where $K_{j,p}$, A and C are as previously defined. Hint: use (2) from the proof of the Lemma.

Proof of Theorem 2. To calculate the variance of $\hat{m}^{(j)}(x)$, using Theorem 1, write

$$\mathrm{var}(h^j \hat{m}^{(j)}(x)/j!) = \frac{\sigma^2}{nhf(x)}[A^{-1}CA^{-1}]_{j+1,j+1} + o\left(\frac{1}{nh}\right)$$

which, by Exercise 3, is equal to

$$\frac{\sigma^2}{nhf(x)} \frac{1}{(j!)^2} \int_{u \in [-x/h,(1-x)/h]} K_{j,p}^2(u)\, du + o\left(\frac{1}{nh}\right).$$

For the bias, by Theorem 1

$$h^{i-1}\ \mathrm{E}[\hat{m}^{(i-1)}(x) - m^{(i-1)}(x)] = h^{p+1} g_1(x)(i-1)![A^{-1}\mathbf{v}]_i$$

$$+h^{p+2}(i-1)!\left\{g_2(x)[A^{-1}\mathbf{w}]_i - \frac{f'(x)}{f(x)} g_1(x)[A^{-1}BA^{-1}\mathbf{v}]_i\right\} + o(h^{p+2}),$$

that is

$$\mathrm{E}[\hat{m}^{(i-1)}(x)] - m^{(i-1)}(x) = h^{p-i+2} g_1(x)(i-1)![A^{-1}\mathbf{v}]_i$$

$$+h^{p-i+3}(i-1)!\left\{g_2(x)[A^{-1}\mathbf{w}]_i - \frac{f'(x)}{f(x)} g_1(x)[A^{-1}BA^{-1}\mathbf{v}]_i\right\} + o(h^{p-i+3}).$$

If K is a probability density symmetric about 0 and if $i+p$ is even then, from Lemma 1

$$\mathrm{E}[\hat{m}^{(i-1)}(x)] - m^{(i-1)}(x) = h^{p-i+2} g_1(x)\ \nu_{p+1}\left(K_{i-1,p}, x, h\right) + o(h^{p-i+2}).$$

If K is a probability density symmetric about 0, if $i+p$ is odd and if $x \in [h, 1-h]$ then, from Lemma 1

$$\mathrm{E}[\hat{m}^{(i-1)}(x)] - m^{(i-1)}(x) = h^{p-i+3}\left[g_2(x)\nu_{p+2}\left(K_{i-1,p}\right) - \frac{f'(x)}{f(x)} g_1(x)(i-1)\nu_{p+1}\left(K_{i-2,p}\right)\right]$$

$$+o(h^{p-i+3}).$$

If $i+p$ is odd and if $x \notin [h, 1-h]$ then

$$\mathrm{E}[\hat{m}^{(i-1)}(x)] - m^{(i-1)}(x) = h^{p-i+2} g_1(x)\ \nu_{p+1}\left(K_{i-1,p}, x, h\right) + o(h^{p-i+2}).$$

The bias results follow by taking $j = i - 1$.

Exercise 4. Suppose that $x \in [h, 1-h]$. Calculate $K_{0,2}$ and $K_{1,2}$. Show directly (from your calculated $K_{i,2}$) that $K_{0,2}$ is a kernel of order 4 and $K_{1,2}$ is a kernel of order $(1,3)$. Plot $K_{0,2}$ and $K_{1,2}$ for $K(u) = \exp(-u^2/2)/\sqrt{2\pi}$. From the plots, why is $K_{0,2}$ good for estimating m? Why is $K_{1,2}$ good for estimating m'?

Proof of the Lemma.

Proof of 1. Let c_{ik} be the determinant of the p by p matrix gotten from A by removing the ith row and kth column. Then the following useful fact follows by expanding the determinant of $M_{(i-1,p)}(u)$ along the ith column.

$$K_{i-1,p}(u) = (i-1)! \sum_k (-1)^{i+k} u^{k-1} \frac{c_{ik}}{\det A} K(u) = (i-1)! \sum_k [A^{-1}]_{ik} u^{k-1} K(u). \tag{2}$$

Thus, for $L = 0, 1, \cdots,$

$$\nu_L\left(K_{i-1,p}, x, h\right) = \int_{u \in [-x/h, (1-x)/h]} u^L K_{i-1,p}(u)\, du$$

$$= (i-1)! \sum_k [A^{-1}]_{ik} \int_{u \in [-x/h, (1-x)/h]} u^{L+k-1}\, K(u)\, du.$$

Setting $L = p+1$ yields the first equality of [1.] and setting $L = p+2$ yields the second.

The following fact will be useful in proving 2, 3 and 4.

$$A_{ij}^{-1} = 0 \quad \text{for } i+j \text{ odd.} \tag{3}$$

To prove this, fix i and j with $i+j$ odd. Let $C \equiv C^{i,j}$ be the p by p matrix gotten from A by deleting the ith row and jth column of A. Then $[A^{-1}]_{ij} = (-1)^{i+j}|C|/|A|$. So it suffices to show that C isn't invertible, that is, that there exists $\mathbf{z}$ not equal to the zero vector with $C\mathbf{z} = 0$. Since A is symmetric, we can consider the case j odd, i even. To make notation simpler, consider $j = 1$. The proof for other j's is similar. Then C is gotten from A

by removing the first column and ith row, and we seek $\mathbf{z}$ such that

$$0 = [C\mathbf{z}]_k = \sum_{s=1}^{p} A_{k,s+1} z_s = \sum_{s=1}^{p} \nu_{k+s-1} z_s \qquad k = 1, \cdots, i{-}1, \cdots, i{+}1, \cdots, p{+}1.$$

Note that the non-zero terms in this sum are s with $k+s$ odd. Let $z_s = 0$, s even. Then we must find z_s s odd, $s = 1, \cdots, p$, not all zero with

$$0 = \sum_{s \text{ odd}, k+s \text{ odd}}^{p} \nu_{k+s-1} z_s \qquad k = 1, \cdots, i-1, \cdots, i+1, \cdots, p+1.$$

This sum is zero if k is odd. So we must find non-zero z_s's, s odd, $s = 1, \cdots, p$, with

$$0 = \sum_{s \text{ odd}}^{p} \nu_{k+s-1} z_s \qquad k \text{ even }, k = 1, \cdots, i-1, \cdots, i+1, \cdots, p+1.$$

Suppose that p is odd. Then we have $(p-1)/2$ linear equations in $(p+1)/2$ unknowns. So there exists a non-zero solution.

Suppose that p is even. Then we have $(p/2) - 1$ linear equations in $p/2$ unknowns. So there exists a non-zero solution.

Proof of 2.

$$[A^{-1}\mathbf{v}]_i = \sum_{j=1}^{p+1} [A^{-1}]_{ij} \nu_{p+j}.$$

Since $\nu_{p+j} = 0$ for $p+j$ odd and $[A^{-1}]_{ij} = 0$ for $i+j$ odd, the sum actually extends over all j with $p+j$ even and $i+j$ even. Suppose that $i+p$ is odd. First consider p even, so that i is odd. Then the sum extends over all j with j even and $i+j$ even. But this is impossible since i is odd, so the sum is zero. Now suppose that p is odd, so that i is even. Then the sum extends over all j with j odd and $i+j$ even. But this is impossible since i is even, so the sum is zero.

Proof of 3.

$$[A^{-1}\mathbf{w}]_i = \sum_{j=1}^{p+1} [A^{-1}]_{ij} \nu_{p+j+1}.$$

Since $\nu_{p+j+1} = 0$ for $p+j$ even and $[A^{-1}]_{ij} = 0$ for $i+j$ odd, the sum actually extends over all j with $p+j$ odd and $i+j$ even. Suppose that $i+p$ is even. First consider p even, so that i is even. Then the sum extends over all j with j odd and $i+j$ even. But this is impossible since i is even, so the sum is zero. Now suppose that p is odd, so that i is odd. Then the sum extends over all j with j even and $i+j$ even. But this is impossible since i is odd, so the sum is zero.

Proof of 4.

$$[A^{-1}BA^{-1}\mathbf{v}]_i = \sum_{j,k,l=1}^{p+1} [A^{-1}]_{ij}\nu_{j+k-1}[A^{-1}]_{kl}\nu_{p+l}.$$

The non-zero terms in the sum are those j, k, and l with $p+l$ even, $k+l$ even, $j+k$ odd, and $i+j$ even.

If p is even and i is even, then the sum extends over those j, k, and l with l even, k even, j odd, and $i+j$ even. This is impossible since i is even, so the sum is zero.

If p is odd and i is odd, then the sum extends over those j, k, and l with l odd, k odd, j even, and $i+j$ even. This is impossible since i is odd, so the sum is zero.

For the remaining cases, we use the fact that $\nu_{j+k-1} = A_{j+1,k}$ for all k and all $j = 1, \cdots, p$ and $\nu_{j+k-1} = A_{j,k+1}$ for all j and all $k = 1, \cdots, p$.

Suppose that p is odd and i is even. Then the non-zero terms in the sum have l odd, k odd, and j even. Since k is odd, the k summation actually runs from $k = 1, \cdots, p$. So the sum is equal to

$$\begin{aligned}\sum_{j,k,l}[A^{-1}]_{ij}A_{j,k+1}[A^{-1}]_{kl}\nu_{p+l} &= \sum_{k,l} I_{i,k+1}[A^{-1}]_{kl}\nu_{p+l} \\ &= \sum_{l}[A^{-1}]_{i-1,l}\nu_{p+l} = \sum_{j,l}[A^{-1}]_{i-1,l}\mathbf{v}_l \\ &= [A^{-1}\mathbf{v}]_{i-1} = \nu_{p+1}(K_{i-2,p})/(i-2)!\end{aligned}$$

by result 1.

Suppose that p is even and i is odd. Then the non-zero terms in the sum have l even, k even, and j odd. Since k is even, the k summation actually runs from $k = 1, \cdots, p$. So the sum is equal to $\nu_{p+1}(K_{i-2,p})$.

4 Minimax Properties of the Local Linear Estimate (optional)

The local linear estimate of $m(x_0)$ is minimax in the following sense. Suppose that the X_i's are iid with density f on $[0,1]$. Fix x_0 an interior point and define

$$\mathcal{S}(C) = \{m : |m(x) - m(x_0) - m'(x_0)(x - x_0)| \le C(x - x_0)^2/2 \text{ for all } x \in [0,1]\}$$

and define $R(n, C)$, the linear minimax risk of estimating $m(x_0)$:

$$R(n, C) = \inf_{\hat{m}(x_0) \text{ linear}} \sup_{m \in \mathcal{S}(C)} \mathrm{E}\left([\hat{m}(x_0) - m(x_0)]^2 | X_1, \cdots, X_n\right).$$

Here, we mean that $\hat{m}(x_0)$ is linear if $\hat{m}(x_0)$ is of the form $\sum_{j=1}^{n} W_j(x_0, X_1, \cdots, X_n) Y_i$.

Suppose that $f(\cdot)$ is continuous at x_0 and $f(x_0) > 0$. Then

$$R(n, C) = \frac{3}{4} 15^{-1/5} \left(\frac{\sqrt{C}\sigma^2}{n f(x_0)}\right) \quad (1 + a_n)$$

where a_n converges to zero in probablity. Furthermore, $\hat{m}(x_0)$, the local linear estimate using the Epanechnikov kernel

$$K(u) = \frac{3}{4}(1 - u^2)_+$$

with bandwidth

$$h = n^{-1/5} \left(\frac{15\sigma^2}{f(x_0) C^2}\right)^{1/5}$$

has

$$\frac{R(n, C)}{\sup_{m \in \mathcal{S}_C} \mathrm{E}\left([\hat{m}(x_0) - m(x_0)]^2 | X_1, \cdots, X_n\right)} \to 1$$

in probability.

Reference: Fan and Gijbels (1996), *Local Polynomial Modelling and its Applications*, Chapman and Hall.

5 Proof of Theorem 1 (optional)

Let

$$\beta_0^* = m(x), \quad \beta_1^* = m'(x), \quad \beta_p^* = m^{(p)}(x)/p!$$

and write

$$\begin{aligned} m(x_i) &= m(x) + m'(x)(x_i - x) + \cdots + m^{(p)}(x)(x_i - x)^p/p! + r(x_i, x) \\ &= \beta_0^* + \beta_1^*(x_i - x) + \cdots + \beta_p^*(x_i - x)^p + r(x_i, x) \end{aligned}$$

where $r(x_i, x)$ is the error in the Taylor series expansion, that is

$$r(x_i, x) = m(x_i) - \sum_0^p \beta_j^*(x_i - x)^j.$$

Let $\mathbf{R} = (r(x_1, x) \cdots, r(x_n, x))^t$ and $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_n)^t$. Then

$$\mathbf{Y} = (m(x_1), \cdots, m(x_n))^t \;+\; \boldsymbol{\epsilon} = X\boldsymbol{\beta}^* + \mathbf{R} + \boldsymbol{\epsilon}$$

and so, using equation (2),

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^* + \left[X^t K_h X\right]^{-1} X^t K_h \left[\mathbf{R} + \boldsymbol{\epsilon}\right].$$

Let H be the $p+1$ by $p+1$ diagonal matrix with

$$H_{ii} = h^{i-1}$$

and write

$$\begin{aligned} E\left[H(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\right] &= H\left[X^t K_h X\right]^{-1} X^t K_h \mathbf{R} \\ &= \left[H^{-1} X^t K_h X H^{-1}\right]^{-1} H^{-1} X^t K_h \mathbf{R} \\ &\equiv U_n^{-1} \mathbf{Z}_n \end{aligned}$$

where

$$U_n = \frac{1}{(n+1)h} H^{-1} X^t K_h X H^{-1}$$

and

$$\mathbf{Z}_n = \frac{1}{(n+1)h} H^{-1} X^t K_h \mathbf{R},$$

and

$$\begin{aligned}
\operatorname{cov}\left[H(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)\right] &= \sigma^2 H\left[X^tK_hX\right]^{-1}X^tK_h^2X\left[X^tK_hX\right]^{-1}H \\
&= \sigma^2\left[H^{-1}X^tK_hXH^{-1}\right]^{-1}H^{-1}X^tK_h^2XH^{-1}\left[H^{-1}X^tK_hXH^{-1}\right]^{-1} \\
&\equiv \frac{\sigma^2}{(n+1)h}\,U_n^{-1}W_nU_n^{-1}
\end{aligned}$$

where

$$W_n = \frac{1}{(n+1)h}H^{-1}X^tK_h^2XH^{-1}.$$

To study the asymptotics of U_n, $\mathbf{Z}_n$, and W_n, we'll use the following fact:

$$\frac{1}{(n+1)h}\sum_{j=1}^{n}\left(\frac{x_j-x}{h}\right)^L K\left(\frac{x_j-x}{h}\right)$$

$$= f(x)\nu_L(K,x,h) + hf'(x)\nu_{L+1}(K,x,h) + o(h) + O\left(\frac{1}{nh^2}\right). \tag{4}$$

Note that, since $nh^3\to\infty$, we can incorporate the $O(1/(nh^2))$ term into the $o(h)$ term. To prove (4), use Lemma 1 of the Nadaraya-Watson handout:

$$\frac{1}{(n+1)h}\sum_j\left(\frac{x_j-x}{h}\right)^L K\left(\frac{x_j-x}{h}\right) = \frac{1}{h}\int\left(\frac{u-x}{h}\right)^L K\left(\frac{u-x}{h}\right)\,f(u)\,du + O\left(\frac{1}{nh^2}\right)$$

$$\begin{aligned}
&= \int_{\mathcal{I}_{x,h}} s^L\;K(s)\;f(x+sh)\;ds + O\left(\frac{1}{nh^2}\right). \\
&= \int_{\mathcal{I}_{x,h}} s^L\;K(s)\;[f(x)+f'(x)sh+o(h)]\;ds + O\left(\frac{1}{nh^2}\right).
\end{aligned}$$

First consider $\mathbf{Z}_n$:

$$[\mathbf{Z}_n]_i = \frac{1}{(n+1)h}\sum_j H_{ii}^{-1}[X^t]_{ij}[K_h]_{jj}R_j$$

$$= \frac{1}{(n+1)h}\sum_j\left(\frac{x_j-x}{h}\right)^{i-1}K\left(\frac{x_j-x}{h}\right)r(x_j,x).$$

But, uniformly in $|x_j - x| \leq h$,

$$r(x_j, x) = \frac{m^{(p+1)}(x)}{(p+1)!}(x_j - x)^{p+1} + \frac{m^{(p+2)}(x)}{(p+2)!}(x_j - x)^{p+2} + o(h^{p+2})$$

so

$$\begin{aligned}
[\mathbf{Z}_n]_i &= \frac{m^{(p+1)}(x)}{(p+1)!} h^{p+1} \frac{1}{(n+1)h} \sum_j \left(\frac{x_j - x}{h}\right)^{i+p} K\left(\frac{x_j - x}{h}\right) \\
&+ \frac{m^{(p+2)}(x)}{(p+2)!}\ h^{p+2} \frac{1}{(n+1)h} \sum_j \left(\frac{x_j - x}{h}\right)^{i+p+1} K\left(\frac{x_j - x}{h}\right) \\
&+ o(h^{p+2}) \frac{1}{(n+1)h} \sum_j \left(\frac{x_j - x}{h}\right)^{i-1} K\left(\frac{x_j - x}{h}\right) \\
&= \frac{m^{(p+1)}(x)}{(p+1)!}\ h^{p+1}\ \left[f(x)\ \nu_{i+p}(K, x, h) + hf'(x)\nu_{i+p+1}(K, x, h) + o(h)\right] \\
&+ \frac{m^{(p+2)}(x)}{(p+2)!}\ h^{p+2}\ \left[f(x)\ \nu_{i+p+1}(K, x, h) + O(h)\right] \\
&\quad + o(h^{p+2}) \\
&= h^{p+1}\ f(x)\ g_1(x)\nu_{i+p}(K, x, h) + h^{p+2}\ f(x)\ g_2(x)\ \nu_{i+p+1}(K, x, h) + o(h^{p+2}).
\end{aligned}$$

Therefore

$$\mathbf{Z}_n = h^{p+1} f(x) g_1(x) \mathbf{v} + h^{p+2} f(x) g_2(x) \mathbf{w} + o(h^{p+2}). \tag{5}$$

Now consider U_n:

$$\begin{aligned}
[U_n]_{ij} &= \frac{1}{(n+1)h} \sum_k H_{ii}^{-1} [X^t]_{ik} [K_h]_{kk} X_{kj} H_{jj}^{-1} \\
&= \frac{1}{(n+1)h} \sum_k \left(\frac{x_k - x}{h}\right)^{i+j-2} K\left(\frac{x_k - x}{h}\right) \\
&= f(x)\ \nu_{i+j-2}\ (K, x, h)\ +\ hf'(x)\ \nu_{i+j-1}\ (K, x, h)\ + o(h) \\
&= f(x)\ A_{ij}\ +\ hf'(x) B_{ij} +\ o(h).
\end{aligned}$$

So

$$U_n = f(x) A + hf'(x) B + o(h). \tag{6}$$

Similarly

$$W_n = f(x)C + O(h) + O\left(\frac{1}{nh^2}\right). \tag{7}$$

Combining (5) and (6) yields

$$E\left[H(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\right] = U_n^{-1}\mathbf{Z}_n$$

$$= \left[f(x)A + hf'(x)B + o(h)\right]^{-1} f(x)\left[h^{p+1}g_1(x)\mathbf{v} + h^{p+2}g_2(x)\mathbf{w} + o(h^{p+2})\right]$$

which, by the lemma below, is equal to

$$\left[\frac{1}{f(x)}A^{-1} - h\frac{f'(x)}{f^2(x)}A^{-1}BA^{-1} + o(h)\right] f(x)\left[h^{p+1}g_1(x)\mathbf{v}+h^{p+2}g_2(x)\mathbf{w}+o(h^{p+2})\right]$$

$$= h^{p+1}g_1(x)A^{-1}\mathbf{v} + h^{p+2}\left[g_2(x)A^{-1}\mathbf{w} - \frac{f'(x)}{f(x)}g_1(x)A^{-1}BA^{-1}\mathbf{v}\right] + o(h^{p+2}).$$

Combining (6) and (7) yields

$$\begin{aligned}\operatorname{cov}\left[H(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\right] &= \frac{\sigma^2}{(n+1)h}\left[f(x)A + O(h)\right]^{-1} \left[f(x)C + O(h)\right]\left[f(x)A + O(h)\right]^{-1} \\ &= \frac{\sigma^2}{nhf(x)}A^{-1}CA^{-1} + o\left(\frac{1}{nh}\right).\end{aligned}$$

Lemma. Let D be an L by L matrix, invertible, and let F be L by L. Suppose that $h_n \to 0$. Then

$$\left[D + h_nF + o(h_n)\right]^{-1} = D^{-1} - h_nD^{-1}FD^{-1} + o(h_n).$$

Proof. We first incorporate the $o(h_n)$ term into F by writing $h_nF + o(h_n) = h_nF_n$ where $F_n \to F$. Then

$$\left[D + h_nF + o(h_n)\right]^{-1} = \left[D + h_nF_n\right]^{-1} = \left[D(I + D^{-1}h_nF_n)\right]^{-1} = \left[I + D^{-1}h_nF_n\right]^{-1} D^{-1}.$$

Thus it suffices to show that

$$\left[I + D^{-1}h_nF_n\right]^{-1} = I - h_nD^{-1}F + o(h_n)$$

that is that

$$\frac{1}{h_n}\left([I + h_n D^{-1} F_n]^{-1} - [I - h_n D^{-1} F]\right) \to 0.$$

The left hand side of the above is equal to

$$\begin{aligned}
\frac{1}{h_n} \quad & [I + h_n D^{-1} F_n]^{-1}\ \left(I - [I + h_n D^{-1} F_n][I - h_n D^{-1} F]\right) \\
&= \frac{1}{h}\ [I + h_n D^{-1} F_n]^{-1}\ \left(I - [I + h_n D^{-1}(F_n - F) - h_n^2 D^{-1} F_n D^{-1} F]\right) \\
&= [I + h_n D^{-1} F_n]^{-1}\ \left(-D^{-1}(F_n - F) + h_n D^{-1} F_n D^{-1} F]\right) \\
& \to 0.
\end{aligned}$$