Consider the regression model $Y_i = r(X_i) + \varepsilon_i, \; i=1,\dots,n$ where ε_i are iid with zero mean and variance $\sigma^2 < \infty$. Assume X_i are also iid and independent of $\{\varepsilon_i\}_{i=1}^n$ (random design).

We define the Nadaraya-Watson estimator as

$$\hat{r}(x) = \frac{\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) Y_i}{\sum_{j=1}^{n} K\left(\frac{x - X_j}{h_n}\right)},$$

where $h_n > 0$ is the <u>bandwidth</u>, and $K(\cdot)$ is a kernel function such that

$$\int_{-\infty}^{\infty} K(x)\,dx = 1, \quad \int_{-\infty}^{\infty} x K(x)\,dx = 0, \quad 0 < \sigma_K^2 = \int_{-\infty}^{\infty} x^2 K(x)\,dx < \infty$$

$$\text{and} \quad \int_{-\infty}^{\infty} K^2(x)\,dx < \infty;$$

<u>Theorem</u>: Let $R(x) = \mathbb{E}\left[\left(\hat{r}(x) - r(x)\right)^2\right]$ and assume X_i's have density f. Furthermore assume both f and r are twice continuously differentiable in $x \in \mathbb{R}$. Then

$$R(x) = \frac{h_n^4}{4} \sigma_K^4 \left(r''(x) + 2r'(x)\frac{f'(x)}{f(x)} \right)^2 + o(h_n^4) + \frac{1}{nh_n}\sigma^2 \left(\int K^2(y)\,dy\right) \frac{1}{f(x)} + o\left(\frac{1}{nh_n}\right) \quad \text{as } \begin{array}{l} h_n \to 0 \\ nh_n \to \infty \end{array}.$$

proof: to ease the notation write $h \equiv h_n$, and define

$$\hat{m}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) Y_i, \qquad \hat{f}(x) = \frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right)$$

Clearly $\hat{r}(x) = \dfrac{\hat{m}(x)}{\hat{f}(x)}$, and we must study the behavior

of $$\hat{r}(x) - r(x) = \frac{\hat{m}(x) - r(x)\hat{f}(x)}{\hat{f}(x)}$$

Since the denominator is random, this is a bit tricky. However, note that

$$\hat{r}(x) - r(x) = \underbrace{\frac{\hat{m}(x) - r(x)\hat{f}(x)}{f(x)}}_{\hat{a}(x)} \; \frac{f(x)}{\hat{f}(x)}$$

Lemma: $\dfrac{f(x)}{\hat{f}(x)} = 1 + o_p(1)$ as $h \to 0$, $nh \to \infty$.

We will prove this later, for now note that it suffices to study $\hat{a}(x)$:

$$\hat{a}(x) = \frac{1}{f(x)} \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x_i - X_i}{h}\right)\left(Y_i - r(x)\right)$$

As we are interested in the quadratic error $\mathbb{E}\left[\hat{\alpha}^2(x)\right]$ we can use the bias-variance decomposition:

Bias:

$$\mathbb{E}\left[\hat{\alpha}(x)\right] = \mathbb{E}\left[\frac{\hat{m}(x) - r(x)\hat{f}(x)}{f(x)}\right] = \frac{1}{f(x)}\mathbb{E}\left[\hat{m}(x)\right] - \frac{r(x)}{f(x)}\mathbb{E}\left[\hat{f}(x)\right]$$

$$\mathbb{E}\left[\hat{m}(x)\right] = \mathbb{E}\left[\frac{1}{nh}\sum K\left(\frac{x - X_i}{h}\right)Y_i\right]$$

$$= \frac{1}{nh}\sum \mathbb{E}\left[\mathbb{E}\left[K\left(\frac{x - X_i}{h}\right)Y_i \,\middle|\, X_i\right]\right]$$

$$= \frac{1}{nh}\sum \mathbb{E}\left[K\left(\frac{x - X_i}{h}\right) r(X_i)\right]$$

$$\overset{iid}{=} \frac{1}{h}\mathbb{E}\left[K\left(\frac{x - X_1}{h}\right) r(X_1)\right]$$

$$= \frac{1}{h}\int K\left(\frac{x - t}{h}\right) r(t) f(t)\, dt$$

$$= \frac{1}{h}\int K\left(\frac{x - t}{h}\right) m(t)\, dt$$

where $m(t) \equiv r(t) f(t)$ is also twice cont. diff. Doing a simple change of variables we get

$$\mathbb{E}\left[\hat{m}(x)\right] = \frac{1}{h}\int K(u)\, m(x + hu)\, h\, du$$

$$= \int K(u)\, m(x + hu)\, du$$

Using a Taylor expansion around $h=0$, we get

$$m(x-hu) = m(x) - m'(x)uh + \frac{1}{2}m''(x)u^2h^2 + u^2 o(h^2) \quad , \; h \to 0$$

Therefore

$$\mathbb{E}\left[\hat{m}(x)\right] = \int K(u)\left(m(x) - m'(x)uh + \frac{1}{2}m''(x)u^2h^2 + u^2 o(h^2)\right)du$$

$$= m(x)\underbrace{\int K(u)\,du}_{=1} - m'(x)h\underbrace{\int uK(u)\,du}_{=0} + \frac{1}{2}m''(x)h^2\underbrace{\int u^2K(u)\,du}_{=\sigma_K^2} + \sigma_K^2\, o(h^2)$$

$$= m(x) + m''(x)\frac{h^2}{2}\sigma_K^2 + o(h^2)$$

Similarly
$$\mathbb{E}\left[\hat{f}(x)\right] = \frac{1}{h}\int K\left(\frac{t-x}{h}\right)f(t)\,dt$$

$$= f(x) + f''(x)\frac{h^2}{2}\sigma_K^2 + o(h^2)$$

Putting all together

$$\mathbb{E}\left[\hat{\alpha}(x)\right] = \frac{1}{f(x)}\left(\cancel{m(x)} + m''(x)\frac{h^2}{2}\sigma_K^2 - \cancel{r(x)f(x)} - r(x)f''(x)\frac{h^2}{2}\sigma_K^2\right) + o(h^2)$$

$$= \frac{h^2}{2f(x)}\sigma_K^2\left(r(x)f''(x) + 2r'(x)f'(x) + r''(x)f(x) - r(x)f''(x)\right) + o(h^2)$$

$$= \frac{h^2}{2}\sigma_K^2\left(r''(x) + 2r'(x)\frac{f'(x)}{f(x)}\right) + o(h^2)$$

since $m''(x) = r''(x) f(x) + 2 r'(x) f'(x) + r(x) f''(x)$.

The square of this term is precisely the firts term is the r.h.s. of the theorem.

Now, for the variance

$$\mathrm{Var}(\hat{a}(x)) = \mathrm{Var}\left(\frac{1}{f(x)} \frac{1}{nh} \sum \underbrace{K\left(\frac{x - X_i}{h}\right)(Y_i - r(x))}_{\text{iid}} \right)$$

$$= \frac{1}{f^2(x)} \frac{1}{(nh)^2} \sum \mathrm{Var}\left(K\left(\frac{x - X_i}{h}\right)(Y_i - r(x)) \right)$$

$$= \frac{1}{f^2(x)} \frac{1}{nh^2} \mathrm{Var}\left(K\left(\frac{x - X_1}{h}\right)(Y_1 - r(x)) \right)$$

$$= \frac{1}{f^2(x)} \frac{1}{nh^2} \mathrm{Var}\left(K\left(\frac{x - X_1}{h}\right)(r(X_1) + \varepsilon_1 - r(x)) \right)$$

$$= \frac{1}{f^2(x)} \frac{1}{nh^2} \mathrm{Var}\left(K\left(\frac{x - X_1}{h}\right)(r(X_1) - r(x) + \varepsilon_1) \right)$$

uncorrelated terms: A, B $\mathbb{E}[AB] = 0$

$$= \frac{1}{f^2(x)} \frac{1}{nh^2} \mathrm{Var}\left(\underbrace{K\left(\frac{x - X_1}{h}\right)(r(X_1) - r(x))}_{A} \right) + \frac{1}{f^2(x)} \frac{1}{nh^2} \mathrm{Var}\left(\underbrace{K\left(\frac{x - X_1}{h}\right)\varepsilon_1}_{B} \right)$$

Now since $\mathbb{E}[B] = 0$ we get

$$Var(B) = \mathbb{E}\left[K^2\left(\frac{x-X_1}{h}\right)\varepsilon_1^2\right] = \sigma^2\,\mathbb{E}\left[K^2\left(\frac{x-X_1}{h}\right)\right]$$

$$= \sigma^2 \int K^2\left(\frac{x-t}{h}\right) f(t)\,dt$$

$$= \sigma^2 h \int K^2(u)\, f(x-hu)\,du$$

$$= \sigma^2 h \int K^2(u)\left(f(x) + o(1)\right) du$$

$$= \sigma^2 h f(x) \int K^2(u)\,du + o(h)$$

$$Var(A) \leq \mathbb{E}\left[K^2\left(\frac{x-X_1}{h}\right)\left(r(X_1)-r(x)\right)^2\right]$$

$$= \int K^2\left(\frac{x-t}{h}\right)\left(r(t)-r(x)\right)^2 f(t)\,dt$$

$$= h\int K^2(u)\left(r(x-hu)-r(x)\right)^2 f(x-hu)\,du$$

$$= h\int K^2(u)\left(o(1)\right) = o(h)$$

Therefore

$$Var(\hat{a}(x)) = \frac{1}{nh}\,\sigma^2\frac{1}{f(x)}\int K^2(u)\,du + o\left(\frac{1}{nh}\right)$$, provided $nh \to \infty$

This is the second term in the rhs of the theorem.

This, together with the lemma above concludes the proof.

<u>Exercise</u>: Prove the lemma, in particular show that $\mathbb{E}[\hat{f}(x)] = f(x) + o(h)$ and $Var(\hat{f}(x)) = \frac{1}{nh}\int K^2(u)\,du + o\left(\frac{1}{nh}\right)$, which implies

$$\hat{f}(x) \xrightarrow{L_2} f(x).$$

The boundary bias issue:

Suppose now the density $f(x)$ has finite support $[a,b]$, namely

$f(x) = g(x)\, 1\{x \in [a,b]\}$, and $g(x)$ is twice cont. differentiable,

For any point in the interior of $[a,b]$ the reasoning above holds, but for $x=a$ or $x=b$ something ~~funny~~ happens:

For simplicity assume $K(x)$ is symmetric, and that it has support $[-1,1]$

$$\mathbb{E}\left[\hat{m}(t)\right] = \frac{1}{h}\int_a^b K\left(\frac{x-t}{h}\right)\underbrace{r(t)g(t)}_{m(t)}\,dt$$

$$= \int_{\frac{x-b}{h}}^{\frac{x-a}{h}} K(u)\, r(x-hu)\, g(x-hu)\, du$$

If $x=a$ we get $\mathbb{E}\left[\hat{m}(t)\right] = m(x)\int_{\frac{x-b}{h}}^{0} K(u)\,du - m'(x)h\int_{\frac{x-b}{h}}^{0} uK(u)\,du + \cdots$

so, for h small enough $= \frac{1}{2}m(x) + m'(x)h\int_{-\infty}^{0} uK(u)\,du + o(h)$

Likewise $\mathbb{E}\left[\hat{f}(x)\right] = \frac{1}{2}f(x) + f'(x)h\int_{-\infty}^{0} uK(u)\,du + o(h)$

So $\mathbb{E}\left[\hat{\alpha}(x)\right] = \frac{1}{f(x)}\left(\frac{1}{2}r(x)g(x) + \left(r'(x)g(x) + r(x)g'(x) - r(x)g'(x)\right)h\int_{-\infty}^{0} uK(u)\,du \;\; ~~- \frac{1}{2}r(x)g(x)~~ \right) + o(h)$

$$= r'(x)h\int_{-\infty}^{0} uK(u)\,du + o(h)$$

So the bias is of the order h near the boundary. It is not hard to see that the variance is still of the order $\frac{1}{nh}$, therefore The estimator is going to suffer from extra bias near the boundaries.

This is a problem, that can be resolved by using local polynomial approaches.