The Nadaraya-Watson Estimator

Derivation of the estimator

We have a random sample of bivariate data $(x_1, Y_1), \ldots, (x_n, Y_n)$.

The Nadaraya-Watson estimator we will be studying in this section is more suitable for a random design. ie. when the data come from a joint pdf $f(x, y)$. The regression model is

$$Y_i = m(x_i) + e_i, \quad i = 1, \ldots, n$$

where $m(\cdot)$ is unknown. The errors $\{\epsilon_i\}$ satisfy

$$E(\epsilon_i) = 0, \quad V(\epsilon_i) = \sigma_\epsilon^2, \quad Cov(\epsilon_i, \epsilon_j) = 0 \text{ for } i \neq j.$$

To derive the estimator note that we can express $m(x)$ in terms of the joint pdf $f(x, y)$ as follows:

$$m(x) = E[Y \mid X = x] = \int y f(y \mid x) dy = \frac{\int y f(x, y) dy}{\int f(x, y) dy}$$

We want to estimate the numerator and denominator separately using kernel estimators. Firstly, for the joint density $f(x, y)$ we use a product kernel density estimator. ie

$$\begin{aligned}
\hat{f}(x, y) &= \frac{1}{n h_x h_y} \sum_{i=1}^n K\left(\frac{x - x_i}{h_x}\right) K\left(\frac{y - y_i}{h_y}\right) \\
&= \frac{1}{n} \sum_{i=1}^n K_{h_x}(x - x_i) K_{h_y}(y - y_i)
\end{aligned}$$

Hence, we have that

$$\int y \hat{f}(x, y) dy = \frac{1}{n} \int y \sum_{i=1}^n K_{h_x}(x - x_i) K_{h_y}(y - y_i)$$

Now, $\int y K_{h_y}(y - y_i) dy = y_i$. Hence, we can write that

$$\int y \hat{f}(x, y) dy = \frac{1}{n} \sum_{i=1}^n K_{h_x}(x - x_i) y_i$$

This is our estimate of the numerator. For the denominator we have

$$\begin{aligned}
\int \hat{f}(x, y) dy &= \frac{1}{n} \sum_{i=1}^n K_{h_x}(x - x_i) \int K_{h_y}(y - y_i) dy \\
&= \frac{1}{n} \sum_{i=1}^n K_{h_x}(x - x_i) \quad \text{since the integral wrt } y \text{ equals one} \\
&= \hat{f}(x)
\end{aligned}$$

Therefore, the Nadaraya-Watson estimate of the unknown regression function is given by

$$\begin{aligned}
\hat{m}(x) &= \frac{\sum_{i=1}^n K_{h_x}(x-x_i)y_i}{\sum_{i=1}^n K_{h_x}(x-x_i)} \\
&= \sum_{i=1}^n W_{h_x}(x, x_i)y_i
\end{aligned}$$

where the weight function $W_{h_x}(x, x_i) = \frac{K_{h_x}(x-x_i)}{\sum_{i=1}^n K_{h_x}(x-x_i)}$. Note that $\sum_{i=1}^n W_{h_x}(x, x_i) = 1$. This kernel regression estimator was first proposed by Nadaraya (1964) and Watson (1964). Note that the estimator is linear in the observations $\{y_i\}$ and is, therefore, a *linear smoother*.

Asymptotic properties

This is complicated by the fact that the estimator is the ratio of two correlated random variables. In the denominator we have that

$$\begin{aligned}
E\hat{f}(x) &\approx f(x) + \frac{h^2}{2}\sigma_K^2 f^{(2)}(x) \\
\text{and } V(\hat{f}(x)) &\approx \frac{R(K)f(x)}{nh}
\end{aligned}$$

(See Section 2 on kernel density estimation)

For the the numerator,

$$\begin{aligned}
E\left[\sum_{i=1}^n K_{h_x}(x-x_i)Y_i\right] &= \int\int v\frac{1}{n}K\left(\frac{x-u}{h_x}\right) f(u,v)dudv \\
&= \int\int vK(s)f(x-hs,v)dsdv \quad (+)
\end{aligned}$$

using the change of variable $s = \frac{x-u}{h_x}$. Now,

$$f(v \mid x-hs) = \frac{f(x-hs,v)}{f(x-hs)}$$

so that $f(x-hs,v) = f(v \mid x-hs)f(x-hs)$. The integral in (+) above is therefore equal to

$$\int\int vK(s)f(v \mid x-hs)f(x-hs)dsdv = \int K(s)f(x-hs)\int vf(v \mid x-hs)dvds$$

$$
\begin{aligned}
&= \int K(s)f(x-hs)m(x-hs)ds \\
&= f(x)m(x) + h_x^2\sigma_K^2[f^{(1)}(x)m^{(1)}(x) + f^{(2)}(x)m(x)/2 + f(x)m^{(2)}(x)/2 + o(h^2)]
\end{aligned}
$$

using Taylor series expansions for $f(x-hs)$ and $m(x-hs)$. Therefore,

$$
\begin{aligned}
E\hat{m}(x) &\approx \frac{E\int \hat{f}(x,y)ydy}{E\hat{f}(x)} \\
&\approx \frac{f(x)[m(x) + h_x^2\sigma_K^2(f^{(1)}m^{(1)}/f + f^{(2)}m/(2f) + m^{(2)}/2)]}{f(x)[1 + h_x^2\sigma_K^2 f^{(2)}/(2f)]} \\
&= m(x) + \frac{h_x^2}{2}\sigma_K^2\left[m^{(2)}(x) + 2m^{(1)}(x)\frac{f^{(1)}(x)}{f(x)}\right]
\end{aligned}
$$

using the approximation that $1 + h^2c)^{-1} \approx (1 - h^2c)$ for small h in the factor in the denominator and multiplying through. Hence, for a random design, the

$$
bias(\hat{m}(x)) \approx \frac{h_x^2}{2}\sigma_K^2\left[m^{(2)}(x) + 2m^{(1)}(x)\frac{f^{(1)}(x)}{f(x)}\right]
$$

However, in the fixed design case the

$$
bias(\hat{m}(x)) \approx \frac{h_x^2}{2}m^{(2)}(x)
$$

When $f^{(1)}(x) = 0$ the bias with a random design equals that with a fixed design. However, the two situations are not identical. The random design has zero probability of being equally-spaced, even when $f(x)$ is the $U(0,1)$ pdf.

The $V(\hat{m}(x))$ can be obtained by using the following approximation for the variance of the ratio of two random variables, N and D:

$$
V\left(\frac{N}{D}\right) \approx \left(\frac{EN}{ED}\right)^2\left[\frac{V(N)}{(EN)^2} + \frac{V(D)}{(ED)^2} - \frac{2Cov(N,D)}{(EN)(ED)}\right]
$$

provided the variance of the ratio exists. This result is based on a first-order Taylor series expansion. Now,

$$
\begin{aligned}
V\left[\frac{1}{n}\sum_{i=1}^{n} K_{h_x}(x-x_i)Y_i\right] &= \frac{1}{n}E[K_{h_x}(x-x_i)Y_i]^2 - O(n^{-1}) \\
&\approx \frac{R(K)f(x)}{nh}[\sigma_\epsilon^2 + m(x)^2]
\end{aligned}
$$

using the facts that $\int v^2 f(v \mid x - hs) = [\sigma_\epsilon^2(x - hs) + m(x - hs)^2]$ and $\sigma_\epsilon^2(x) = \sigma_\epsilon^2$ for all x. (ie. a constant).

Also,

$$V(\hat{f}(x)) \approx \frac{R(K)f(x)}{nh}$$

Finally,

$$\begin{aligned} Cov\left[\frac{1}{n}\sum_{i=1}^{n} K_{h_x}(x - x_i)Y_i, \frac{1}{n}\sum_{i=1}^{n} K_{h_x}(x - x_i)\right] &= \frac{1}{n}E[K_{h_x}(x - x_i)^2 Y_i] - O(n^{-1}) \\ &\approx \frac{R(K)f(x)m(x)}{nh} \end{aligned}$$

Substituting into the approximation formula gives

$$V(\hat{m}(x)) \approx \frac{R(K)\sigma_\epsilon^2}{nhf(x)}$$

The variance of $\hat{m}(x)$ involves terms relating to the error variance σ_ϵ^2 and the relative amount of data through $f(x)$.

We can use the above point-wise bias and variance results to construct an expression for the AMSE of $\hat{m}(x)$ which is as follows:

$$AMSE(\hat{m}(x)) \approx \frac{h_x^4}{4}\sigma_K^4\left[m^{(2)}(x) + 2m^{(1)}(x)\frac{f^{(1)}(x)}{f(x)}\right]^2 + \frac{R(K)\sigma_\epsilon^2}{nhf(x)}$$