Solution to Exercise 15.25

⋆15.25 For a given choice of bandwidth, the expectation of the estimate $\hat{f}_h(x)$ of (15.61) is h^{-1} times the expectation of the random variable $k\big((x-X)/h\big)$, where X denotes the random variable of which the x_t are IID realizations. Assume that k is symmetric about the origin. Show that the bias of $\hat{f}_h(x)$ is independent of the sample size n and roughly proportional to h^2 for small h. More formally, this means that the bias is $O(h^2)$ as $h \to 0$.

Show also that the variance of $\hat{f}_h(x)$ is of order $(nh)^{-1}$ as $n \to \infty$ and $h \to 0$. Why do these facts imply that the bandwidth h that minimizes the expectation of the squared error of $\hat{f}_h(x)$ must be of order $n^{-1/5}$ as $n \to \infty$?

The expectation of the kernel density estimator (15.61) is

$$\begin{aligned}\mathrm{E}\big(\hat{f}_h(x)\big) &= h^{-1}\,\mathrm{E}\Big(k\big((x-X)/h\big)\Big)\\ &= h^{-1}\!\int_{-\infty}^{\infty} k\big((x-X)/h\big)\, f(X)\,dX\\ &= \int_{-\infty}^{\infty} k(W) f(x-hW)\,dW,\end{aligned} \tag{S15.49}$$

where $W \equiv (x-X)/h$. A second-order Taylor expansion of $f(x-hW)$ around the point $h=0$ yields

$$f(x-hW) = f(x) - hWf'(x) + \tfrac{1}{2}h^2W^2f''(x) + O(h^3). \tag{S15.50}$$

Substituting this into (S15.49), we find that $\mathrm{E}\big(\hat{f}_h(x)\big)$ is the sum of three integrals and a term of order h^3. The first integral is

$$\int_{-\infty}^{\infty} k(W) f(x)\,dW = f(x)\int_{-\infty}^{\infty} k(W)\,dW = f(x).$$

The second integral is the negative of

$$\int_{-\infty}^{\infty} k(W) hWf'(x)\,dW = hf'(x)\int_{-\infty}^{\infty} Wk(W)\,dW = 0,$$

because the mean of W is 0. The third integral is

$$\int_{-\infty}^{\infty} k(W)\tfrac{1}{2}h^2W^2f''(x) = \tfrac{1}{2}h^2f''(x)\int_{-\infty}^{\infty} W^2k(W)\,dW = \tfrac{1}{2}h^2f''(x),$$

because the variance of W is 1. Thus we conclude that

$$\mathrm{E}\big(\hat{f}_h(x)\big) - f(x) = \tfrac{1}{2}h^2f''(x) + O(h^3), \tag{S15.51}$$

which is evidently $O(h^2)$.

Because the kernel density estimator (15.61) is an average of n IID random variables, its variance is n^{-1} times the common variance of these variables. That is,

$$\operatorname{Var}\big(\hat{f}_h(x)\big) = \frac{1}{nh^2}\operatorname{Var}\big(k(x - X/h)\big). \tag{S15.52}$$

The variance on the right-hand side above can be computed as follows. We have

$$\begin{aligned}\operatorname{Var}\big(k(x - X/h)\big) &= \int_{-\infty}^{\infty} k^2\big((x - X)/h\big) f(X)\,dX - \mathrm{E}^2\big(k(x - X)/h\big)\\ &= h\int_{-\infty}^{\infty} k^2(W) f(x - hW)\,dW - \mathrm{E}^2\big(k(x - X)/h\big).\end{aligned} \tag{S15.53}$$

If we substitute the second-order Taylor expansion (S15.50) into the integral in expression (S15.53), we once again obtain three integrals and a term of order h^3. The first integral is

$$\int_{-\infty}^{\infty} k^2(W) f(x)\,dW = f(x)\int_{-\infty}^{\infty} k^2(W)\,dW. \tag{S15.54}$$

The second integral is the negative of

$$h\int_{-\infty}^{\infty} k^2(W) W f'(x)\,dW = hf'(x)\int_{-\infty}^{\infty} W k^2(W)\,dW = 0,$$

because the kernel function $k(W)$ is symmetric. The third integral is

$$\int_{-\infty}^{\infty} k^2(W)\tfrac{1}{2}h^2 W^2 f''(x) = \tfrac{1}{2}h^2 f''(x)\int_{-\infty}^{\infty} W^2 k^2(W)\,dW.$$

Since $\mathrm{E}(k(x - X/h))$ is h times $\mathrm{E}(\hat{f}_h(x))$, we find from (S15.51) and (S15.53), along with the above results, that

$$\operatorname{Var}\big(k(x - X/h)\big) = hf(x)\int_{-\infty}^{\infty} k^2(W)\,dW - h^2 f^2(x) + O(h^3),$$

and so, from (S15.52), we see that

$$\operatorname{Var}\big(\hat{f}_h(x)\big) = (nh)^{-1}\bigg(f(x)\int_{-\infty}^{\infty} k^2(W)\,dW - O(h)\bigg). \tag{S15.55}$$

Since we are considering the limit where $h \to 0$, the term above that is $O(h)$ is asymptotically negligible relative to the other term, which is of order $(nh)^{-1}$, as we wished to show.

The squared error is the variance plus the square of the bias. Therefore, the results (S15.51) and (S15.55) imply that the approximate squared error is

$O(h^4) + O(n^{-1})O(h^{-1})$. If we differentiate this with respect to h and set the derivative to 0 to find the optimal value of h, we find that

$$O(n^{-1})O(h^{-2}) + O(h^3) = 0.$$

Multiplying by h^2 and rearranging, this becomes

$$O(n^{-1}) = -O(h^5),$$

from which we conclude that $h = O(n^{-1/5})$ if it is chosen optimally, as we were required to show.