2 Univariate Density Estimation

2.1 Numerical derivative of distribution function

$$
\begin{aligned}
\widehat{f}(x) =& \frac{\widehat{F}(x+h) - \widehat{F}(x-h)}{2h} \\
=& \frac{\sum_{i=1}^{n} \mathbf{1}\left\{x_i \leq x+h\right\} - \sum_{i=1}^{n} \mathbf{1}\left\{x_i \leq x-h\right\}}{2nh} \\
=& \frac{\sum_{i=1}^{n} \left[\mathbf{1}\left\{x_i \leq x+h\right\} - \mathbf{1}\left\{x_i \leq x-h\right\}\right]}{2nh} \\
=& \frac{\sum_{i=1}^{n} \mathbf{1}\left\{\left|\frac{x_i - x}{h}\right| \leq 1\right\}}{2nh}
\end{aligned}
$$

2.2 Density integrates to unity

To ensure that our density estimator integrates to 1, we can see that

$$
\begin{aligned}
\int\limits_{\mathcal{S}(x)} \widehat{f}(x)dx = \int\limits_{\mathcal{S}(x)} (nh)^{-1} \sum_{i=1}^{n} k\left(\left|\frac{x_i - x}{h}\right|\right) dx \\
=& (nh)^{-1} \sum_{i=1}^{n} \int\limits_{\mathcal{S}(x)} k\left(\left|\frac{x_i - x}{h}\right|\right) dx \\
=& n^{-1} \sum_{i=1}^{n} \int\limits_{\mathcal{S}(x)} k(t)dt = n^{-1} \sum_{i=1}^{n} 1 = n^{-1}n = 1.
\end{aligned}
$$

The second equality flips the orders of integration and summation while the third equality uses the change of variable $\left|\frac{x_i - x}{h}\right| = t$, implying that $dx/h = dt$, and we have assumed that $\int\limits_{\mathcal{S}(x)} k(\psi)d\psi = 1$. So, for our density estimator to be an actual density, we require that the kernel we use is nonnegative everywhere and integrates to one. Notice that these are not restrictive assumptions since the user controls the kernel employed in a statistical analysis. Additionally, as we will see later, other properties of our estimators (such as bias and variance) also depend on the properties of the kernel used.

Second moment of equivalent kernel integrates to unity

$$\begin{aligned}\kappa_2(\tilde{k}) =& \int t^2 \kappa_2(k)^{1/2} k(\kappa_2(k)^{1/2} t) dt \\ =& \int \kappa_2(k)^{-1} \psi^2 k(\psi) d\psi = \kappa_2(k)^{-1} \int \psi^2 k(\psi) d\psi \\ =& \kappa_2(k)^{-1} \kappa_2(k) = 1,\end{aligned}$$

where we have used the change of variable $\psi = \kappa_2(k)^{1/2} t$, implying that $t = \kappa_2(k)^{-1/2}\psi$, $t^2 = \kappa_2(k)^{-1}\psi^2$ and $d\psi = \kappa_2(k)^{1/2} dt$.

2.3 Bias of density estimator

We first derive the expected value of the estimator of the density

$$\begin{aligned}E\left[\widehat{f}(x)\right] =& E\left[\frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x_i - x}{h}\right)\right] = \frac{1}{nh}\sum_{i=1}^{n} E\left[k\left(\frac{x_i - x}{h}\right)\right] \\ =& \frac{1}{nh}\sum_{i=1}^{n} E\left[k\left(\frac{x_1 - x}{h}\right)\right] = h^{-1} E\left[k\left(\frac{x_1 - x}{h}\right)\right].\end{aligned}$$

The first equality follows by definition of the kernel density estimator, the second by the linearity of the expectation operator, the third by the fact that the data have an identical distribution and the last by cancelling out the n^{-1} with the n identical summands. Continuing we have

$$\begin{aligned}E\left[\widehat{f}(x)\right] =& h^{-1} E\left[k\left(\frac{x_1 - x}{h}\right)\right] = h^{-1}\int k\left(\frac{z - x}{h}\right) f(z) dz \\ =& \int k(\psi) f(x + h\psi) d\psi. \qquad (1)\end{aligned}$$

We have used the change of variable $\psi = (z - x)/h$ which yields $z = x + h\psi$ and $dz = h^{-1} d\psi$. In general, this integral is not analytically solvable unless $f(z)$ is specified. To get around this issue we can take a Taylor expansion of $f(x + h\psi)$ around x. We take a 2^{nd} order expansion since we have assumed that $f(\psi)$ possesses two continuous derivatives. This gives us

$$f(x + h\psi) \approx f(x) + (h\psi) f^{(1)}(x) + \frac{(h\psi)^2}{2} f^{(2)}(x) + o(h^2).$$

We can plug this approximation into our last derivation for the expectation of $\widehat{f}(x)$ to obtain

$$\begin{aligned}
E\left[\widehat{f}(x)\right] =& \int k(\psi)f(x+h\psi)d\psi \\
\approx& \int \left(f(x) + (h\psi)f^{(1)}(x) + \frac{(h\psi)^2}{2}f^{(2)} + o(h^2) \right) k(\psi)d\psi \\
=& f(x)\int k(\psi)d\psi + hf^{(1)}(x)\int \psi k(\psi)d\psi + \frac{h^2}{2}f^{(2)}(x)\int \psi^2 k(\psi)d\psi + o(h^2) \\
=& f(x)\kappa_0(k) + hf^{(1)}(x)\kappa_1(k) + \frac{h^2}{2}f^{(2)}(x)\kappa_2(k) + o(h^2) \\
=& f(x) + \frac{h^2}{2}f^{(2)}(x)\kappa_2(k) + o(h^2).
\end{aligned} \tag{2}$$

We can see the gains from using a kernel which has $\kappa_2(k) = 1$ from this derivation. Also, note that we used $\kappa_0(k) = 1$ and $\kappa_1(k) = 0$ in the last equality. Having the expectation of our kernel density estimator will allow us to calculate the bias, which is given as

$$Bias(\widehat{f}(x)) = E[\widehat{f}(x)] - f(x) = \frac{h^2}{2}f^{(2)}(x)\kappa_2(k) + o(h^2).$$

2.4 Variance of density estimator

Having obtained $E[\widehat{f}(x)]$, we can now follow similar steps to determine the variance of our density estimator. Once the variance is obtained we can calculate the mean square error (AMSE) of our density estimator. To derive the variance estimator we have

$$\begin{aligned}
Var\left[\widehat{f}(x)\right] =& Var\left[\frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x_i - x}{h}\right)\right] = \frac{1}{n^2h^2}Var\left[\sum_{i=1}^{n} k\left(\frac{x_i - x}{h}\right)\right] \\
=& \frac{1}{n^2h^2}\sum_{i=1}^{n} Var\left[k\left(\frac{x_i - x}{h}\right)\right] = \frac{1}{n^2h^2}\sum_{i=1}^{n} Var\left[k\left(\frac{x_1 - x}{h}\right)\right] \\
=& \frac{1}{nh^2}Var\left[k\left(\frac{x_1 - x}{h}\right)\right],
\end{aligned}$$

where the first equality follows by the definition of the kernel density estimator, the second by the properties of the variance operator, the third, by the fact that the data are independent, the fourth by the identical distributiveness of the data and the last by the summation canceling with an n^{-1}. From this we have

$$\begin{aligned}Var\left[\widehat{f}(x)\right] =& \frac{1}{nh^2}Var\left[k\left(\frac{x_1-x}{h}\right)\right]\\ =& \frac{1}{nh^2}\left(E\left[k\left(\frac{x_1-x}{h}\right)^2\right] - E\left[k\left(\frac{x_1-x}{h}\right)\right]^2\right).\end{aligned}$$

The second term was already calculated and appears in equation (2). We need to calculate the first term to obtain a solution for the variance of our kernel density estimator. Following similar steps used to derive the bias it follows that

$$E\left[k\left(\frac{x_1-x}{h}\right)^2\right] = \int k\left(\frac{z-x}{h}\right)^2 f(z)dz = h\int k(\psi)^2 f(x+h\psi)d\psi.$$

Again, we have used the change of variable $\psi = (z-x)/h$ which yields $z = x + h\psi$ and $dz = h^{-1}d\psi$. Taking a 2^{nd} order expansion around x,

$$f(x+h\psi) \approx f(x) + (h\psi)f^{(1)}(x) + \frac{(h\psi)^2}{2}f^{(2)}(x) + o(h^2),$$

provides

$$\begin{aligned}E\left[k\left(\frac{x_1-x}{h}\right)^2\right] \approx& h\int\left(f(x) + (h\psi)f^{(1)}(x) + \frac{(h\psi)^2}{2}f^{(2)}(x) + o(h^2)\right)k(\psi)^2 d\psi\\ =& hf(x)\int k(\psi)^2 d\psi + h^2 f^{(1)}(x)\int \psi k(\psi)^2 d\psi\\ &+ \frac{h^3}{2}f^{(2)}(x)\int \psi^2 k(\psi)^2 d\psi + o(h^3)\\ =& hf(x)\int k(\psi)^2 d\psi + o(h^2).\end{aligned}$$

The first equality replaced $f(x+h\psi)$ with its Taylor expansion approximation while the last equality collected all of the terms of order h^2 or smaller. Plugging in this approximation for the expectation of $k\left(\frac{x_1-x}{h}\right)^2$ as well as our expectation of $k\left(\frac{x_1-x}{h}\right)$ we have that the variance of our kernel density estimator is

$$
\begin{aligned}
Var\left[\widehat{f}(x)\right] =& \frac{1}{nh^2}\left(E\left[k\left(\frac{x_1-x}{h}\right)^2\right] - E\left[k\left(\frac{x_1-x}{h}\right)\right]^2\right)\\
=& \frac{1}{nh^2}\left(hf(x)\int k(\psi)^2 d\psi + o(h^2)\right.\\
&\left. -h^2\left(f(x)+\frac{h^2}{2}f^{(2)}(x)\kappa_2(k)+o(h^2)\right)^2\right)\\
=& \frac{1}{nh}f(x)\int k(\psi)^2 d\psi + o(n^{-1})\\
&- n^{-1}\left(f(x)+\frac{h^2}{2}f^{(2)}(x)\kappa_2(k)+o(h^2)\right)^2\\
=& \frac{1}{nh}f(x)R(k)+o(n^{-1}).
\end{aligned}
$$

2.5 Calculus of variation problem

Recall that we can always scale the kernel such that $\kappa_2(\tilde{k}) = 1$. That is,

$$
\begin{aligned}
R(\tilde{k}) =& \int \tilde{k}(\psi)d\psi = \int \kappa_2(k)k(\kappa_2(k)^{1/2}t)^2 d\psi\\
=& \int \kappa_2(k)^{1/2}k(w)^2 dw = \kappa_2(k)^{1/2}\int k(w)^2 dw\\
=& \kappa_2(k)^{1/2}R(k).
\end{aligned}
$$

This yields,

$$
\left(R(\tilde{k})^4\right)^{1/5} = \left(R(k)^4\kappa_2^2(k)\right)^{1/5}.
$$

Thus, the problem of optimal kernel selection boils down to the calculus of variation problem:

$$
\underset{k(\psi)}{min} : R(k) \text{ s.t. } : \kappa_0(k) = 1, : \kappa_1(k) = 0, : \kappa_2(k) = 1.
$$

2.6 Kernel efficiency

$$
\begin{aligned}
\mathrm{Eff}(k_\varrho) &= \left[\frac{AMISE_{opt}(\widehat{f}(x);k_\varrho)}{AMISE_{opt}(\widehat{f}(x);k_1)}\right]^{5/4} \\
&= \left[\frac{\frac{5}{4}\left(R(k_\varrho)^4\kappa_2^2(k_\varrho)\right)^{1/5}R(f^{(2)})n^{-4/5}}{\frac{5}{4}\left(R(k_1)^4\kappa_2^2(k_1)\right)^{1/5}R(f^{(2)})n^{-4/5}}\right]^{5/4} \\
&= \left[\frac{\left(R(k_\varrho)^4\kappa_2^2(k_\varrho)\right)^{1/5}}{\left(R(k_1)^4\kappa_2^2(k_1)\right)^{1/5}}\right]^{5/4} = \frac{R(k_\varrho)\kappa_2(k_\varrho)^{1/2}}{R(k_1)}.
\end{aligned}
$$

2.7 Equivalent samples

$$
\begin{aligned}
AMISE_{opt}(\widehat{f}(x);k_1) &= \frac{5}{4}\left(R(k_1)^4\kappa_2^2(k_1)\right)^{1/5}R(f^{(2)})^{1/5}n_1^{-4/5} \\
&= \frac{5}{4}\left(R(k_\varrho)^4\kappa_2^2(k_\varrho)\right)^{1/5}R(f^{(2)})^{1/5}n_\varrho^{-4/5} \\
&= AMISE_{opt}(\widehat{f}(x);k_\varrho),
\end{aligned}
$$

which implies

$$
\begin{aligned}
n_\varrho &= \left[\frac{\left(R(k_1)^4\kappa_2^2(k_1)\right)^{1/5}}{\left(R(k_\varrho)^4\kappa_2^2(k_\varrho)\right)^{1/5}}n_1^{-4/5}\right]^{-5/4} \\
&= \left[\frac{\left(R(k_\varrho)^4\kappa_2^2(k_\varrho)\right)^{1/5}}{\left(R(k_1)^4\kappa_2^2(k_1)\right)^{1/5}}\right]^{5/4} n_1 \\
&= \mathrm{Eff}(k_\varrho)\cdot n_1.
\end{aligned}
$$

Canonical kernel AMISE

Kernels of the form

$$k_\delta(\psi) = \delta^{-1}k(\psi/\delta)$$

are considered and δ is determined by the balancing in AMISE. This can be accomplished through integration by substitution as follows:

$$\int k_\delta(t)^2 dt \equiv \left[\int t^2 k_\delta(t)dt\right]^2 \Rightarrow \int \delta^{-2}k(t/\delta)^2 dt \equiv \left[\int (t^2/\delta)k(t/\delta)dt\right]^2.$$

Using the substitution $\psi = t/\delta$, $d\psi = dt/\delta$ and replacing t with $\delta\psi$ we have

$$\delta^{-1}\int k(\psi)^2 d\psi \equiv \left[\int \delta^2\psi^2 k(\psi)d\psi\right]^2 \Rightarrow \delta^{-1}\int k(\psi)^2 d\psi \equiv \delta^4\left[\int \psi^2 k(\psi)d\psi\right]^2.$$

Using our previous notation are rearranging we are left with

$$\delta^5 \equiv R(k)\kappa_2^{-2}(k) \Rightarrow \delta = R(k)^{1/5}\kappa_2(k)^{-2/5}.$$

Using this canonical kernel we can recalculate AMISE as

$$\begin{aligned} AMISE(\widehat{f}(x)) =& \frac{h^4\kappa_2(k_\delta)^2}{4}R(f^{(2)}) - \frac{R(k_\delta)}{nh} \\ =& \frac{h^4\left[\int(\psi^2/\delta)k(\psi/\delta)d\psi\right]^2}{4}R(f^{(2)}) - \frac{\int \delta^{-2}k(\psi/\delta)^2 d\psi}{nh}. \end{aligned}$$

Using a change of variables we condense this as

$$\begin{aligned} AMISE(\widehat{f}(x)) =& \frac{h^4\delta^4\left[\int t^2k(t)dt\right]^2}{4}R(f^{(2)}) - \frac{\int k(t)^2dt}{nh\delta} \\ =& \frac{h^4R(k)^{4/5}\kappa_2(k)^{-8/5}\kappa_2(k)^2}{4}R(f^{(2)}) - \frac{R(k)}{nhR(k)^{1/5}\kappa_2(k)^{-2/5}} \\ =& \frac{h^4R(k)^{4/5}\kappa_2(k)^{2/5}}{4}R(f^{(2)}) - \frac{R(k)^{4/5}\kappa_2(k)^{2/5}}{nh} \\ =& \left(R(k)^4\kappa_2^2(k)\right)^{1/5}\left[(nh)^{-1} + (h^4/4)R(f^{(2)})\right]. \end{aligned}$$

2.8 Optimal bandwidth for AMSE

$$\begin{aligned}\frac{dMSE(\widehat{f}(x))}{dh} =& \kappa_2^2(k) f^{(2)}(x)^2 h^3 - \frac{f(x)R(k)}{nh^2} = 0\\ &\Rightarrow \kappa_2^2(k) f^{(2)}(x)^2 h_{opt}^5 - \frac{f(x)R(k)}{n} = 0\\ &\Rightarrow h_{opt}^5 = \frac{f(x)R(k)}{n\kappa_2^2(k) f^{(2)}(x)^2}\\ &\Rightarrow h_{opt} = \left[\frac{f(x)R(k)}{\kappa_2^2(k) f^{(2)}(x)^2}\right]^{1/5} n^{-1/5}.\end{aligned}$$

Straightforward algebra gives us, ignoring the terms of $O(h^4 + (nh)^{-1})$,

$$\begin{aligned}AMSE_{opt}(\widehat{f}(x)) =& n^{-1} f(x)R(k)h_{opt}^{-1} + \frac{\kappa_2^2(k)}{4} f^{(2)}(x)^2 h_{opt}^4 +\\ =& n^{-1} f(x)R(k)\left[\frac{f(x)R(k)}{\kappa_2^2(k) f^{(2)}(x)^2}\right]^{-1/5} n^{1/5}\\ &+ \frac{\kappa_2^2(k)}{4} f^{(2)}(x)^2 \left[\frac{f(x)R(k)}{\kappa_2^2(k) f^{(2)}(x)^2}\right]^{4/5} n^{-4/5}\\ =& \left(f(x)R(k)\right)^{4/5}\left(\kappa_2^2(k) f^{(2)}(x)^2\right)^{1/5} n^{-4/5}\\ &+ \frac{1}{4}\left(f(x)R(k)\right)^{4/5}\left(\kappa_2^2(k) f^{(2)}(x)^2\right)^{1/5} n^{-4/5}\\ =& \frac{5}{4}\left(f(x)^4 R^4(k)\kappa_2^2(k) f^{(2)}(x)^2\right)^{1/5} n^{-4/5}.\end{aligned}$$

2.9 Optimal bandwidth for AMISE

$$\begin{aligned}\frac{dMISE(\widehat{f}(x))}{dh} =& \kappa_2^2(k) R(f^{(2)}) h^3 - \frac{R(k)}{nh^2} = 0\\ &\Rightarrow \kappa_2^2(k) R(f^{(2)}) h_{opt}^5 - \frac{R(k)}{n} = 0\\ &\Rightarrow h_{opt}^5 = \frac{R(k)}{n\kappa_2^2(k) R(f^{(2)})}\\ &\Rightarrow h_{opt} = \left[\frac{R(k)}{\kappa_2^2(k) R(f^{(2)})}\right]^{1/5} n^{-1/5}.\end{aligned}$$

Again, straightforward algebra shows us that the AMISE for the optimal bandwidth is

$$\begin{aligned}
AMISE_{opt}(\widehat{f}(x)) =& n^{-1}R(k)h_{opt}^{-1} + \frac{\kappa_2^2(k)}{4}R(f^{(2)})h_{opt}^4 + \\
=& n^{-1}R(k)\left[\frac{R(k)}{\kappa_2^2(k)R(f^{(2)})}\right]^{-1/5} n^{1/5} \\
& + \frac{\kappa_2^2(k)}{4}R(f^{(2)})\left[\frac{R(k)}{\kappa_2^2(k)R(f^{(2)})}\right]^{4/5} n^{-4/5} \\
=& (R(k))^{4/5}\left(\kappa_2^2(k)R(f^{(2)})\right)^{1/5} n^{-4/5} \\
& + \frac{1}{4}(R(k))^{4/5}\left(\kappa_2^2(k)R(f^{(2)})\right)^{1/5} n^{-4/5} \\
=& \frac{5}{4}\left(R^4(k)\kappa_2^2(k)R(f^{(2)})\right)^{1/5} n^{-4/5}.
\end{aligned}$$

2.10 Calculation of $R(f^{(2)})$

Before calculating $R(f^{(2)})$ we note that

$$\sigma^{-1}\phi^{(2)}(\psi/\sigma) = \sigma^{-5}\left(\psi^2 - \sigma^2\right)\phi(\psi/\sigma).$$

Using this information our difficulty factor is calculated as

$$\begin{aligned}
R(f^{(2)}) =& \sigma^{-10}\int\left(\psi^2-\sigma^2\right)^2\phi(\psi/\sigma)^2 d\psi \\
=& \sigma^{-10}\left(\int\psi^4\phi(\psi/\sigma)^2 d\psi - 2\sigma^2\int\psi^2\phi(\psi/\sigma)^2 d\psi + \sigma^4\int\phi(\psi/\sigma)^2 d\psi\right)
\end{aligned}$$

These integrals are easily calculated if we note that

$$\phi(\psi/\sigma)^2 = (1/\sqrt{2\pi})\phi(\sqrt{2}\psi).$$

With this we have

$$
\begin{aligned}
R(f^{(2)}) =& (1/\sqrt{2\pi})\sigma^{-10}\left(\int \psi^4\phi(\sqrt{2}\psi/\sigma)d\psi - 2\sigma^2\int \psi^2\phi(\sqrt{2}\psi/\sigma)d\psi + \sigma^4\int \phi(\sqrt{2}\psi/\sigma)d\psi\right) \\
=& (1/\sqrt{2\pi})\sigma^{-10}\left((\sigma/\sqrt{2})\int \psi^4\sqrt{2}\sigma^{-1}\phi(\sqrt{2}\psi/\sigma)d\psi - 2\sigma^3/\sqrt{2}\int \psi^2\sqrt{2}\sigma^{-1}\phi(\sqrt{2}\psi/\sigma)d\psi \right. \\
& \left. + \sigma^5/\sqrt{2}\int \sqrt{2}\sigma^{-1}\phi(\sqrt{2}\psi/\sigma)d\psi\right) \\
=& (1/\sqrt{2\pi})\sigma^{-10}\left((\sigma/\sqrt{2})\mu_4 - 2\sigma^3/\sqrt{2}\mu_2 + \sigma^5/\sqrt{2}\right),
\end{aligned}
$$

where μ_2 and μ_4 are the 2^{nd} and 4th moments of a normally distributed random variable with variance $\sigma^2/2$. In this case $\mu_2 = \sigma^2/2$ while $\mu_4 = (4!\sigma^4)/(2!2^6) = 3\sigma^4/4$. Combining results we have

$$
R(f^{(2)}) = (1/\sqrt{2\pi})\sigma^{-10}\left(3\sigma^5/(4\sqrt{2}) - \sigma^5/\sqrt{2} + \sigma^5/\sqrt{2}\right) = \frac{3}{8\sqrt{\pi}\sigma^5}.
$$

2.11 Unknown roughness bias

From our discussion on density derivative estimation we have that

$$
\begin{aligned}
R(\widehat{f}^{(2)}) &= \int \left[\frac{1}{nh^3}\sum_{i=1}^{n} k^{(2)}\left(\frac{x-x_i}{h}\right)\right]^2 dx \\
&= \int \frac{1}{n^2h^6}\sum_{i=1}^{n}\sum_{j=1}^{n} k^{(2)}\left(\frac{x-x_i}{h}\right)k^{(2)}\left(\frac{x-x_j}{h}\right)dx \\
&= \int \frac{1}{n^2h^6}\sum_{i=1}^{n} k^{(2)}\left(\frac{x-x_i}{h}\right)^2 dx \\
&\quad + \int \frac{1}{n^2h^6}\sum_{i=1}^{n}\sum_{j=1}^{n} k^{(2)}\left(\frac{x-x_i}{h}\right)^2 k^{(2)}\left(\frac{x-x_j}{h}\right)dx. \qquad (3)
\end{aligned}
$$

Taking expectations we have

$$
\begin{aligned}
E\left[R(\widehat{f}^{(2)})\right] =& E\left[\int \frac{1}{n^2h^6}\sum_{i=1}^{n} k^{(2)}\left(\frac{x-x_i}{h}\right)^2 dx\right] \\
&+ E\left[\int \frac{1}{n^2h^6}\sum_{i=1}^{n}\sum_{j=1}^{n} k^{(2)}\left(\frac{x-x_i}{h}\right)^2 k^{(2)}\left(\frac{x-x_j}{h}\right)\right] dx \\
=& \int \frac{1}{n^2h^6}\sum_{i=1}^{n} E\left[k^{(2)}\left(\frac{x-x_i}{h}\right)^2\right] dx \\
&+ \int \frac{1}{n^2h^6}\sum_{i=1}^{n}\sum_{j=1}^{n} E\left[k^{(2)}\left(\frac{x-x_i}{h}\right) k^{(2)}\left(\frac{x-x_j}{h}\right) dx\right] \\
=& \int \frac{1}{nh^6} E\left[k^{(2)}\left(\frac{x-x_1}{h}\right)^2\right] dx \\
&+ \int \frac{n-1}{nh^6} E\left[k^{(2)}\left(\frac{x-x_1}{h}\right) k^{(2)}\left(\frac{x-x_2}{h}\right)\right] dx \\
=& \int \frac{1}{nh^6} E\left[k^{(2)}\left(\frac{x-x_1}{h}\right)^2 dx\right] \qquad (4) \\
&+ \int \frac{n-1}{nh^6} E\left[k^{(2)}\left(\frac{x-x_1}{h}\right)\right] E\left[k^{(2)}\left(\frac{x-x_2}{h}\right)\right] dx. \qquad (5)
\end{aligned}
$$

We analyze each of the terms of the expectation separately. Beginning with (4), we have

$$
\begin{aligned}
\int \frac{1}{nh^6} E\left[k^{(2)}\left(\frac{x-x_1}{h}\right)^2 dx\right] &= \int\int \frac{1}{nh^6} k^{(2)}\left(\frac{x-z}{h}\right)^2 f(z)dzdx \\
&= \int\int \frac{1}{nh^5} k^{(2)}(\psi)^2 f(x-h\psi)d\psi dx \\
&= \frac{1}{nh^5}\int k^{(2)}(\psi)^2 \int f(x-h\psi)d\psi dx \\
&= \frac{1}{nh^5}\int k^{(2)}(\psi)^2 dx = \frac{R(k^{(2)})}{nh^5}, \qquad (6)
\end{aligned}
$$

where we used the change of variables $\psi = (x-z)/h$, $d\psi = dz/h$ and $z = x - h\psi$. Moving to the analysis of (5) we obtain

$$
\int \frac{n-1}{nh^6} E\left[k^{(2)}\left(\frac{x-x_1}{h}\right)\right] E\left[k^{(2)}\left(\frac{x-x_2}{h}\right)\right] dx = \int \frac{n-1}{nh^6}\left[\int k^{(2)}\left(\frac{x-z}{h}\right) f(z)dz\right]^2 dx.
$$

This follows via the identical distribution of the data. Continuing we have

$$
\begin{aligned}
&= \int \frac{n-1}{n} \left[\int \frac{1}{h^3} k^{(2)} \left(\frac{x-z}{h} \right) f(z) dz \right]^2 dx \\
&= \int \frac{n-1}{n} \left[\int \frac{1}{h} k \left(\frac{x-z}{h} \right) f^{(2)}(z) dz \right]^2 dx \\
&= \int \frac{n-1}{n} \left[\int k(\psi) f^{(2)}(x - h\psi) d\psi \right]^2 dx \\
&= \int \frac{n-1}{n} \left[\int k(\psi) \left(f^{(2)}(x) - (h\psi) f^{(1)}(x) + (h\psi)^2 f^{(2)}(x) + o_p(h^2) \right) d\psi \right]^2 dx \\
&= \frac{n-1}{n} \int \left[\left[\int k(\psi) f^{(2)}(x) d\psi \right]^2 + \left[\int o_p(h^2) d\psi \right] \right] dx \\
&= \frac{n-1}{n} \int \left[f^{(2)}(x) \int k(\psi) d\psi \right]^2 dx + O_p(h^2) = \frac{n-1}{n} \int f^{(2)}(x)^2 dx + O_p(h^2), \qquad (7)
\end{aligned}
$$

where the first equality follows by definition, the second via application of integration by parts twice (as was done in our discussion of the bias and variance of the kernel density derivative), the third via a change of variables $\psi = (x-z)/h$, $d\psi = dz/h$ and $z = x - h\psi$, the fourth by a second-order Taylor expansion around x, and the last few by rearranging terms and combining things of smaller order, the last equality following by the integration of the kernel to 1.

Combining our results in (6) and (7) we obtain

$$
E\left[R(\widehat{f}^{(2)}) \right] = \frac{R(k^{(2)})}{nh^5} + \frac{n-1}{n} R(f^{(2)}) + O_p(h^2),
$$

which for large n we have $E\left[R(\widehat{f}^{(2)}) \right] = \frac{R(k^{(2)})}{nh^5} + R(f^{(2)}) + O_p(h^2)$. This result shows that using the roughness of a kernel density derivative to estimate the unknown roughness produces a positively biased estimate (Since $R(k^{(2)}) > 0$) which is asymptotically nonvanishing if we use the optimal order of h. That is, we know that $h_{opt} = cn^{-1/5}$ and upon substitution into our expectation we achieve

$$
\frac{R(k^{(2)})}{nh_{opt}^5} = \frac{R(k^{(2)})}{n(cn^{-1/5})^5} = \frac{R(k^{(2)})}{c^{1/5} n n^{-1}} = R(k^{(2)}) c^{-1/5}.
$$

2.12 Expectation of average leave-one-out estimator

To see this consider

$$E\left[\bar{f}_{-i}(x)\right] = n^{-1}\sum_{j=1}^{n} E\left[\widehat{f}_{-j}(x)\right] = E\left[\widehat{f}_{-1}(x)\right],$$

since $E\left[\widehat{f}_{-i}(x)\right] = E\left[\widehat{f}_{-j}(x)\right]$, for all i and j.

Now to see why this quantity is important, consider,

$$\begin{aligned}
E\left[\widehat{f}_{-1}(x)\right] &= E\left[E\left[\widehat{f}_{-1}(x)\middle| x_2, x_3, \ldots, x_n\right]\right] \\
&= E\left[\int \widehat{f}_{-1}(x)f(x)dx\right] = \int E\left[\widehat{f}_{-1}(x)\right] f(x)dx \\
&= \int E\left[\widehat{f}(x)\right] f(x)dx = E\left[\int \widehat{f}(x)f(x)dx\right]
\end{aligned}$$

where we have exchanged expectation and integration and used the fact that $E\left[\widehat{f}_{-1}(x)\right] = E\left[\widehat{f}(x)\right]$ due to independence of this quantity on the sample size. Thus, $\widehat{f}_{-i}(x)$ is an unbiased estimator of $\int \widehat{f}(x)f(x)dx$. Given that we want to avoid obtaining bandwidths which are too heavily influenced by a particular observation and to abrogate choice of observation to leave out in the construction of $\widehat{f}_{-i}(x)$, we use $\bar{f}_{-i}(x)$ which is also an unbiased estimator of$\int \widehat{f}(x)f(x)dx$.

2.13 Derivation of LSCV

Turning our attention to the first term in (**??**), we have

$$\begin{aligned}
\int \widehat{f}(x)^2 dx &= \int \left(\frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x_i - x}{h}\right)\right)^2 dx \\
&= \frac{1}{n^2h^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\int k\left(\frac{x_i - x}{h}\right) k\left(\frac{x_j - x}{h}\right) dx.
\end{aligned}$$

Note that the integral in the double summation is, with an appropriate change of variables, the definition of a convolution. That is, the convolution of a function with itself is $\bar{g}(x) =$

$\int g(\psi)g(x-\psi)d\psi$.[1] If we make the change of variable $\psi = \frac{x_i - x}{h}$, we have

$$\begin{aligned}\frac{1}{h}\int k\left(\frac{x_i - x}{h}\right)k\left(\frac{x_j - x}{h}\right)dx =& \frac{1}{h}\int k\left(\psi\right)k\left(\psi - \frac{x_j - x_i}{h}\right)d\psi \\ =& \bar{k}\left(\frac{x_j - x_i}{h}\right).\end{aligned}$$

If we combine the leave-one-out estimator with what we just derived, we can see that the LSCV criterion can be written as

$$\begin{aligned} LSCV(h) &= \int \widehat{f}(x)^2 dx - 2\int \widehat{f}(x)f(x)dx \\ &= \frac{1}{n^2 h}\sum_{i=1}^{n}\sum_{j=1}^{n}\bar{k}\left(\frac{x_i - x_j}{h}\right) - \frac{2}{n(n-1)}\sum_{i=1}^{n}\widehat{f}_{-i}(x_i) \\ &= \frac{1}{n^2 h}\sum_{i=1}^{n}\sum_{j=1}^{n}\bar{k}\left(\frac{x_i - x_j}{h}\right) - \frac{2}{n\left(n-1\right)h}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}k\left(\frac{x_j - x}{h}\right)\end{aligned}$$

In general, no exact closed form solution exists for the convolution kernels within the s-class of kernels. The convolution kernel for the Gaussian kernel is easy to find since the convolution of two normal densities is itself a normal density. For the distinct polynomial class, each kernel, $k_s(\psi)$, has support on $[-1, 1]$ and so it follows that $\bar{k}(x)$ has support on $[-2, 2]$. For example, if $x = 1.25$ and $\psi = 0.5$, then $x - \psi = 0.75 < 1$. For $x \geq 0$ we have

$$\bar{k}_s(x) = \int_{x-1}^{1} k_s(\psi)k_s(x-\psi)d\psi.$$

This integral can be calculated using algebraic software. We use Maxima for the convolution kernels presented in the equations below. The convolution kernels are symmetric about 0.

[1] When $g(x)$ is symmetric, like our kernel, we have that $\bar{g}(x) = \int g(u)g(x-u)du = \int g(u)g(u-x)du$.

We can find $\bar{k}(x)$, for $x < 0$, by using the relation $\bar{k}(x) = \bar{k}(-x)$:

$$
\begin{aligned}
\bar{k}_0(x) =& \frac{1}{4}(2 - |x|)\mathbf{1}\{|x| \leq 2\} \\
\bar{k}_1(x) =& \frac{3}{160}(2 - |x|)^3(x^2 + 6|x| + 4)\mathbf{1}\{|x| \leq 2\} \\
\bar{k}_2(x) =& \frac{5}{3584}(2 - |x|)^5(x^4 + 10|x|^3 + 36x^2 + 40|x| + 16)\mathbf{1}\{|x| \leq 2\} \\
\bar{k}_3(x) =& \frac{35}{1757184}(2 - |x|)^7(5x^6 + 70|x|^5 + 404x^4 + 1176|x|^3 \\
& \qquad + 1616x^2 + 1120|x| + 320)\mathbf{1}\{|x| \leq 2\} \\
\bar{k}_\phi(x) =& \frac{1}{\sqrt{4\pi}} e^{-x^2/4}.
\end{aligned}
$$

2.14 Convolution kernels and LSCV

$$
\begin{aligned}
LSCV(h) =& \frac{1}{n^2 h} \sum_{i=1}^{n} \sum_{j=1}^{n} \bar{k}\left(\frac{x_i - x_j}{h}\right) - \frac{2}{(n-1)n} \sum_{i=1}^{n} \widehat{f}_{-i}(x_i) \\
=& \frac{1}{n^2 h} \sum_{i=1}^{n} \sum_{j=1}^{n} \bar{k}\left(\frac{x_i - x_j}{h}\right) - \frac{2}{(n-1)nh} \sum_{i=1}^{n} \sum_{j=1}^{n} k\left(\frac{x_j - x_i}{h}\right) + \frac{2}{(n-1)nh} k(0) \\
=& \frac{1}{n^2 h} \sum_{i=1}^{n} \sum_{j=1}^{n} \left[\bar{k}\left(\frac{x_i - x_j}{h}\right) - \frac{2n}{n-1} k\left(\frac{x_j - x_i}{h}\right) + \frac{2n}{n-1} k(0)\right] \\
=& \frac{1}{n^2 h} \sum_{i=1}^{n} \sum_{j=1}^{n} \left[\tilde{k}\left(\frac{x_i - x_j}{h}\right)\right].
\end{aligned}
$$

2.15 Expected value of LCV

$$\begin{aligned}
E[LCV(h)] =& E[n^{-1}\sum_{i=1}^{n}\log\widehat{f}_{-i}(x_i)] = n^{-1}\sum_{i=1}^{n}E[\log\widehat{f}_{-i}(x_i)] = E[\log\widehat{f}_{-n}(x_n)] \\
=& E\left[E[\log\widehat{f}_{-n}(x)|x_1,\ldots,x_{n-1}]\right] = E\left[\int\log\{\widehat{f}_{-n}(x)\}f(x)dx\right] \\
=& E\left[\int\log\{\widehat{f}(x)\}f(x)dx\right] \\
=& E\left[\int\log\{\widehat{f}(x)\}f(x)dx\right] - \int f(x)\log\{f(x)\}dx + \int f(x)\log\{f(x)\}dx \\
=& -E\left[\int f(x)\log\{f(x)\}dx - \int\log\{\widehat{f}(x)\}f(x)dx\right] + \int f(x)\log\{f(x)\}dx \\
=& -E\left[\int\log\{f(x)/\widehat{f}(x)\}f(x)dx\right] + \int f(x)\log\{f(x)\}dx.
\end{aligned}$$

2.16 Integrating φ^{-1}

$$\begin{aligned}
\varphi^{-1} =& 2\int_0^\infty e^{-0.5(\log(1+x))^2}dx = 2\int_0^\infty (1+x)e^{-0.5(\log(1+x))^2}\frac{dx}{1+x} \\
=& 2\int_0^\infty e^{\psi}e^{-0.5\psi^2}d\psi = \sqrt{2e\pi}(erf(1/\sqrt{2})+1)
\end{aligned}$$

where we used the change of variable $\psi = log(1+x)$, $d\psi = dx/(1+x)$ and $erf(\psi)$ is the error function. Noting that $\Phi(x) = 0.5*[1+erf(x/\sqrt{2})]$, we obtain $erf(1/\sqrt{2}) = 2\Phi(1)-1$, which upon substitution yields, $\varphi^{-1} = 2\Phi(1)\sqrt{2\pi e}$.

LCV optimal bandwidth

$$\begin{aligned}\frac{-C_1 n/2}{(nh)^{3/2}} + 4C_2h^3 = 0 \Rrightarrow & \frac{-C_1 n^{-1/2}/2}{h^{3/2}} + 4C_2h^3 = 0\\ \Rrightarrow & -C_1n^{-1/2}/2 + 4C_2h^{9/2} = 0\\ \Rrightarrow & C_1n^{-1/2}/2 = 4C_2h^{9/2}\\ \Rrightarrow & \frac{C_1}{8C_2}n^{-1/2} = h^{9/2}\\ \Rrightarrow & \left(Cn^{-1/2}\right)^{2/9} = h\\ \Rrightarrow & \tilde{C}n^{-1/9} = h_{opt}.\end{aligned}$$

2.17 Bias of derivative estimator

Before proceeding we mention here that in order to evaluate the integrals which will appear in both the bias and variance of our density derivative estimator we will need to use integration by parts. Recall that integration by parts is defined via

$$\int \psi dv = \psi v - \int v d\psi. \tag{8}$$

If the limits of integration are fixed then the term ψv becomes $\psi v\Big|_a^b$. Focusing on the bias first, we have

$$E\left[\widehat{f}^{(r)}(x)\right] = E\left[\frac{1}{h^{1+r}}k^{(r)}\left(\frac{x_1-x}{h}\right)\right] = \frac{1}{h^{1+r}}\int k^{(r)}\left(\frac{z-x}{h}\right)f(z)dz.$$

To use integration by parts we note that

$$\int \frac{1}{h}k^{(r)}\left(\frac{z-x}{h}\right)dz = -k^{(r-1)}\left(\frac{z-x}{h}\right).$$

Using our definition of integration by parts in (8) we have

$$\begin{aligned}\psi =& f(z)/h^r; & d\psi =& \frac{f^{(1)}(z)}{h^r}dz,\\ v =& -k^{(r-1)}\left(\frac{z-x}{h}\right); & dv =& \frac{1}{h}k^{(r)}\left(\frac{z-x}{h}\right)dz,\end{aligned}$$

which allows us to write our expectation as

$$E\left[\widehat{f}^{(r)}(x)\right] = \frac{-f(z)}{h^r}k^{(r-1)}\left(\frac{z-x}{h}\right)\Big|_{-\infty}^{\infty} + \int \frac{1}{h^r}k^{(r-1)}\left(\frac{z-x}{h}\right)f^{(1)}(z)dz. \tag{9}$$

The first term is zero for the types of kernels commonly employed in empirical work as well as those satisfying the common conditions imposed in theoretical work. Essentially we assume that as $\psi \to \pm\infty$, $k(\psi) \to 0$ in a monotonically decreasing fashion (the further away you are the less weight that you receive) suggesting that the derivatives take on this property as well. Thus,

$$E\left[\widehat{f}^{(r)}(x)\right] = \frac{1}{h^{1+r}}\int k^{(r)}\left(\frac{z-x}{h}\right)f(z)dz = \int \frac{1}{h^r}k^{(r-1)}\left(\frac{z-x}{h}\right)f^{(1)}(z)dz. \tag{10}$$

We can repeat this process of integration by parts another $r-1$ times to arrive at

$$E\left[\widehat{f}^{(r)}(x)\right] = \int \frac{1}{h}k\left(\frac{z-x}{h}\right)f^{(r)}(z)dz. \tag{11}$$

Using the change of variables $\psi = (z-x)/h$ and $d\psi = dz/h$ we have

$$E\left[\widehat{f}^{(r)}(x)\right] = \int k(\psi)f^{(r)}(x+h\psi)d\psi. \tag{12}$$

This is identical to (1) except that we have $f^{(r)}(x+h\psi)$ instead of $f(x+h\psi)$. If we assume that $f^{(r)}(x)$ possesses two continuous derivatives (or that $f(x)$ possesses $r+2$ continuous derivatives) we can follow the exact same procedure to obtain $E\left[\widehat{f}^{(r)}(x)\right]$. That is, taking a second-order Taylor expansion around x we have

$$f^{(r)}(x+h\psi) \approx f^{(r)}(x) + (h\psi)f^{(r+1)}(x) + \frac{(h\psi)^2}{2}f^{(r+2)}(x) + o(h^2).$$

We can plug this approximation into our last derivation for the expectation of $\widehat{f}^{(r)}(x)$ to obtain

$$\begin{aligned}
E\left[\widehat{f}^{(r)}(x)\right] =& \int k(\psi) f^{(r)}(x+h\psi)d\psi \\
\approx& \int \left(f^{(r)}(x) + (h\psi) f^{(r+1)}(x) + \frac{(h\psi)^2}{2} f^{(r+2)} + o(h^2) \right) k(\psi) d\psi \\
=& f^{(r)}(x) \int k(\psi)d\psi + h f^{(r+1)}(x) \int \psi k(\psi) d\psi + \frac{h^2}{2} f^{(2)}(x) \int \psi^2 k(\psi) d\psi + o(h^2) \\
=& f^{(r)}(x)\kappa_0(k) + h f^{(r+1)}(x)\kappa_1(k) + \frac{h^2}{2} f^{(r+2)}(x)\kappa_2(k) + o(h^2) \\
=& f^{(r)}(x) + \frac{h^2}{2} f^{(r+2)}(x)\kappa_2(k) + o(h^2). \qquad (13)
\end{aligned}$$

We used $\kappa_0(k) = 1$ and $\kappa_1(k) = 0$ in the last equality, which underscores the importance of using integration by parts to take the integral so that $k(\psi)$ appears instead of $k^{(r)}(\psi)$ so that we can invoke the properties of the kernel already placed on the problem. Having this expectation allows us to calculate the bias, which is given as

$$Bias(\widehat{f}^{(r)}(x)) = E[\widehat{f}^{(r)}(x)] - f^{(r)}(x) = \frac{h^2}{2} f^{(r+2)}(x)\kappa_2(k) + o(h^2).$$

2.18 Variance of derivative estimator

$$\begin{aligned}
Var\left[\widehat{f}^{(r)}(x)\right] =& \frac{1}{n^2h^{2+2r}} \sum_{i=1}^{n} Var\left[k^{(r)}\left(\frac{x_i - x}{h}\right)\right] \\
=& \frac{1}{nh^{2+2r}} \left(E\left[k^{(r)}\left(\frac{x_1 - x}{h}\right)^2\right] - \left(E\left[k^{(r)}\left(\frac{x_1 - x}{h}\right)\right]\right)^2 \right).
\end{aligned}$$

The second piece requires a slight bit of manipulation to put it into the form that makes it look like that of the bias. Continuing we have

$$\begin{aligned}
Var\left[\widehat{f}^{(r)}(x)\right] = \frac{1}{nh^{2+2r}} \Bigg(& \int k^{(r)}\left(\frac{z-x}{h}\right)^2 f(z)dz \\
& - (h^{1+r})^2 \left(\int \frac{1}{h^{1+r}} k^{(r)}\left(\frac{z-x}{h}\right) f(z)dz \right)^2 \Bigg).
\end{aligned}$$

The second integral is identical to that we came across when analyzing the bias of $\widehat{f}^{(r)}(x)$.

However, note that this second integral has a h^{2+2r} in front of it which cancels the h^{2+2r} that appears in the denominator of our variance. These cancel so that the second half of our variance can be written as

$$\frac{-1}{n}\left(\left(f^{(r)}(x)+\frac{h^2}{2}f^{(r+2)}(x)\kappa_2(k)+o(h^2)\right)^2\right)=\frac{-1}{n}\left(f^{(r)}(x)^2+o(h^2)\right)=o(n^{-1}).$$

Thus, our variance of $\widehat{f}^{(r)}(x)$ simplifies to

$$Var\left[\widehat{f}^{(r)}(x)\right]=\frac{1}{nh^{2+2r}}\left(\int k^{(r)}\left(\frac{z-x}{h}\right)^2 f(z)dz\right)+o(n^{-1}).$$

Again, making the change of variable $\psi=(z-x)/h$ and $d\psi=dz/h$ we have

$$Var\left[\widehat{f}^{(r)}(x)\right]=\frac{1}{nh^{1+2r}}\left(\int k^{(r)}(\psi)^2 f(x+h\psi)d\psi\right)+o(n^{-1}).$$

As we did with the variance of the kernel density estimator we take a second-order Taylor expansion around x to obtain

$$\begin{aligned}
Var\left[\widehat{f}^{(r)}(x)\right]=&\frac{1}{nh^{1+2r}}\left(\int k^{(r)}(\psi)^2\left(f(x)+(h\psi)f^{(1)}(x)+\frac{(h\psi)^2}{2}f^{(2)}+o(h^2)\right)d\psi\right)+o(n^{-1})\\
\approx&\frac{f(x)R(k^{(r)})}{nh^{1+2r}}+o\left(n^{-1}h^{-(1+2r)}\right)+o(n^{-1})\\
\approx&\frac{f(x)R(k^{(r)})}{nh^{1+2r}}+o\left(n^{-1}h^{-(1+2r)}\right).
\end{aligned}$$

2.19 Optimal bandwidth for derivative estimator

$$
\begin{aligned}
\frac{dAMISE(\widehat{f}^{(r)})}{dh} =& h^3 R(f^{(r+2)})\kappa_2^2(k) - (1+2r)\frac{R(k^{(r)})}{nh^{2+2r}} \equiv 0 \\
\rightarrow & h^3 R(f^{(r+2)})\kappa_2^2(k) = (1+2r)\frac{R(k^{(r)})}{nh^{2+2r}} \\
\rightarrow & h^{5+2r} R(f^{(r+2)})\kappa_2^2(k) = (1+2r)n^{-1}R(k^{(r)}) \\
\rightarrow & h_{opt}^{5+2r} = \frac{(1+2r)R(k^{(r)})}{R(f^{(r+2)})\kappa_2^2(k)} n^{-1} \\
\Rrightarrow & h_{opt} = \left[\frac{(1+2r)R(k^{(r)})}{R(f^{(r+2)})\kappa_2^2(k)}\right]^{1/(5+2r)} n^{-1/(5+2r)} \\
& = D_r(k)D_r(f)n^{-1/(5+2r)}
\end{aligned}
$$

where $D_r(k) = \left(\frac{(1+2r)R(k^{(r)})}{\kappa_2^2(k)}\right)^{1/(5+2r)}$ and $D_r(f) = R\left(f^{(r+2)}\right)^{-1/(5+2r)}$.

Using this bandwidth we see that the AMISE is

$$
\begin{aligned}
AMISE_{opt} =& \frac{h_{opt}^4 \kappa_r^2(k)}{4} R(f^{(r+2)}) + \frac{R(k^{(r)})}{n h_{opt}^{1+2r}} \\
=& \frac{\left(D_r(k) D_r(f) n^{-1/(5+2r)}\right)^4 \kappa_r^2(k)}{4} R(f^{(r+2)}) \\
&+ \frac{R(k^{(r)})}{n \left(D_r(k) D_r(f) n^{-1/(5+2r)}\right)^{1+2r}} \\
=& \frac{\left(\left(\frac{(1+2r)R(k^{(r)})}{\kappa_2^2(k)}\right)^{1/(5+2r)} R\left(f^{(r+2)}\right)^{-1/(5+2r)} n^{-1/(5+2r)}\right)^4 \kappa_2^2(k^{(r)})}{4} R(f^{(r+2)}) \\
&+ \frac{R(k^{(r)})}{n \left(\left(\frac{(1+2r)R(k^{(r)})}{\kappa_2^2(k)}\right)^{1/(5+2r)} R\left(f^{(r+2)}\right)^{-1/(5+2r)} n^{-1/(5+2r)}\right)^{1+2r}} \\
=& \frac{\left((1+2r)^{4/(5+2r)} R(k^{(r)})^{4/(5+2r)} \kappa_2^2(k)^{-4/(5+2r)} R\left(f^{(r+2)}\right)^{-4/(5+2r)} n^{-4/(5+2r)}\right) \kappa_2^2(k)}{4} R(f^{(r+2)}) \\
&+ \frac{R(k^{(r)}) n^{(1+2r)/(5+2r)} \kappa_2^2(k)^{(1+2r)/(5+2r)} R\left(f^{(r+2)}\right)^{(1+2r)/(5+2r)}}{n(1+2r)^{(1+2r)/(5+2r)} R(k^{(r)})^{(1+2r)/(5+2r)}} \\
=& \frac{\left((1+2r)^{4/(5+2r)} R(k^{(r)})^{4/(5+2r)} \kappa_2^2(k)^{(1+2r)/(5+2r)} R\left(f^{(r+2)}\right)^{(1+2r)/(5+2r)} n^{-4/(5+2r)}\right)}{4} \\
&+ \frac{R(k^{(r)})^{4/(5+2r)} n^{-4/(5+2r)} \kappa_2^2(k)^{(1+2r)/(5+2r)} R\left(f^{(r+2)}\right)^{(1+2r)/(5+2r)}}{(1+2r)^{(1+2r)/(5+2r)}} \\
=& \Bigg[\Bigg(\frac{(1+2r)^{4/(5+2r)}}{4} \\
&\quad + \frac{1}{(1+2r)^{(1+2r)/(5+2r)}}\Bigg) \left(R(k^{(r)})^4 \left(\kappa_2^2(k) R\left(f^{(r+2)}\right)\right)^{(1+2r)}\right)^{1/(5+2r)}\Bigg] n^{-4/(5+2r)} \\
=& \Bigg[\Bigg(\frac{(1+2r)^{4/(5+2r)}(1+2r)^{(1+2r)/(5+2r)}}{4(1+2r)^{(1+2r)/(5+2r)}} \\
&\quad + \frac{4}{4(1+2r)^{(1+2r)/(5+2r)}}\Bigg) \left(R(k^{(r)})^4 \left(\kappa_2^2(k) R\left(f^{(r+2)}\right)\right)^{(1+2r)}\right)^{1/(5+2r)}\Bigg] n^{-4/(5+2r)} \\
=& \left[\left(\frac{1+2r+4}{4(1+2r)^{(1+2r)/(5+2r)}}\right) \left(R(k^{(r)})^4 \left(\kappa_2^2(k) R\left(f^{(r+2)}\right)\right)^{(1+2r)}\right)^{1/(5+2r)}\right] n^{-4/(5+2r)} \\
=& \frac{5+2r}{4} \left[R(k^{(r)})^4 \left(\frac{\kappa_2^2(k) R\left(f^{(r+2)}\right)}{(1+2r)}\right)^{(1+2r)}\right]^{1/(5+2r)} n^{-4/(5+2r)} \\
=& \frac{5+2r}{4} \left[C_r(k^{(r)}) \left(\frac{R\left(f^{(r+2)}\right)}{(1+2r)}\right)^{(1+2r)}\right]^{1/(5+2r)} n^{-4/(5+2r)},
\end{aligned}
$$

where $C_r(k^{(r)}) = R(k^{(r)})^4 \left(\kappa_2^2(k)\right)^{(1+2r)}$

2.20 Kernel efficiency for derivative estimators

$$
\begin{aligned}
\mathrm{Eff}(k_\varrho, r) &= \left[\frac{AMISE_{opt}(\widehat{f}(x); k_\varrho)}{AMISE_{opt}(\widehat{f}(x); k_{r+1})}\right]^{(5+2r)/4} \\
&= \left[\frac{\frac{5+2r}{4}\left[C_r(k^{(r)})\left(\frac{R\left(f^{(r+2)}\right)}{(1+2r)}\right)^{(1+2r)}\right]^{1/(5+2r)} n^{-4/(5+2r)}}{\frac{5+2r}{4}\left[C_r(k_{r+1}^{(r)})\left(\frac{R\left(f^{(r+2)}\right)}{(1+2r)}\right)^{(1+2r)}\right]^{1/(5+2r)} n^{-4/(5+2r)}}\right]^{(5+2r)/4} \\
&= \left[\left(\frac{C(k_\varrho)}{C(k_{r+1})}\right)^{1/(5+2r)}\right]^{(5+2r)/4} \\
&= \left[\frac{R(k_\varrho^{(r)})^4 \left(\kappa_2^2(k_\varrho)\right)^{(1+2r)}}{R(k_{r+1}^{(r)})^4 \left(\kappa_2^2(k_{r+1})\right)^{(1+2r)}}\right]^{1/4} \\
&= \frac{R(k_\varrho^{(r)})}{R(k_{r+1}^{(r)})}\left(\frac{\kappa_2^2(k_\varrho)}{\kappa_2^2(k_{r+1})}\right)^{(1+2r)/4} .
\end{aligned}
$$