Kernel density estimation

Kernel density estimation.

a. Recall from the notes the general results on the bias (up to $O(h^4)$) and variance (up to order $O(\frac{h}{n})$) of a "simplified kernel" density estimator.

b. **(4 marks)** Prove that the following kernel is a "simplified kernel" (in the sense of the notes).

$$\begin{aligned}
\int_{-1}^{1} K(w)dw &= \int_{-1}^{1} \frac{15}{32}(1-w^2)(3-7w^2)dw \\
&= \int_{-1}^{1} \frac{15}{32}(7w^4-10w^2+3)dw \\
&= \frac{15}{32}(\frac{7w^5}{5}-\frac{10w^3}{3}+3w)|_{-1}^{1} \\
&= \frac{15}{32}[(\frac{7}{5}-\frac{10}{3}+3)-(\frac{-7}{5}+\frac{10}{3}-3)] \\
&= 1
\end{aligned}$$

$$\begin{aligned}
\int_{-1}^{1} wK(w)dw &= \int_{-1}^{1} \frac{15}{32}w(1-w^2)(3-7w^2)dw \\
&= \int_{-1}^{1} \frac{15}{32}(7w^5-10w^3+3w)dw \\
&= \frac{15}{32}(\frac{7w^6}{6}-\frac{10w^4}{4}+\frac{3w^2}{2})|_{-1}^{1} \\
&= \frac{15}{32}[(\frac{7}{6}-\frac{10}{4}+\frac{3}{2})-(\frac{7}{6}-\frac{10}{4}+\frac{3}{2})] \\
&= 0
\end{aligned}$$

$$\begin{aligned}
\int_{-1}^{1} w^2K(w)dw &= \int_{-1}^{1} \frac{15}{32}w^2(1-w^2)(3-7w^2)dw \\
&= \int_{-1}^{1} \frac{15}{32}(7w^6-10w^4+3w^2)dw \\
&= \frac{15}{32}(\frac{7w^7}{7}-\frac{10w^5}{5}+\frac{3w^3}{3})|_{-1}^{1} \\
&= \frac{15}{32}(w^7-2w^5+w^3)|_{-1}^{1} \\
&= \frac{15}{32}[(1-2+1)-(-1+2-1)] \\
&= 0 = \sigma_k^2 \ (0 \le \sigma_k^2 < \infty)
\end{aligned}$$

Therefore, $K(w)$ is a simplified kernel.

ii. **(8 marks)** Determine the (approximate) mean squared error (from the slides) of $\tilde{f}_K(x)$ for the above kernel K and for arbitrary $f(x)$.

$$\begin{aligned}
Bias[\tilde{f}_k(x)] &= \frac{1}{2}\sigma_k^2 h^2 f''(x) + O(h^3) \\
&= 0 + O(h^3) \ (as\ \sigma_k^2 = 0) \\
&= O(h^3)
\end{aligned}$$

$$\begin{aligned}
Var[\tilde{f}_k(x)] &= \frac{1}{nh} f(x) \int_{-1}^{1} K^2(w)dw - \frac{1}{n}[f(x)]^2 + O(\frac{h^2}{n}) \\
&= \frac{1}{nh} f(x) \frac{5}{4} - \frac{1}{n}[f(x)]^2 + O(\frac{h^2}{n}) \\
&= \frac{5}{4nh} - \frac{1}{n}[f(x)]^2 + O(\frac{h^2}{n})
\end{aligned}$$

$$\begin{aligned}
MSE &= Bias^2 + Variance \\
&= Bias[\tilde{f}_k(x)]^2 + Var[\tilde{f}_k(x)] \\
&= O(h^6) + \frac{5}{4nh} - \frac{1}{n}[f(x)]^2 + O(\frac{h^2}{n}) \\
&= \frac{5}{4nh} - \frac{1}{n}[f(x)]^2 + O(h^6) \text{ Since } O(h^6) \text{ converges slower than } O(\frac{h^2}{n})
\end{aligned}$$

iii. **(5 marks)** Determine the case for the above K when the true underlying density $f(x)$ is $N(0, 1)$.

$$\begin{aligned}
f(x) &= \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \\
MSE &= \frac{5}{4nh} - \frac{1}{n}[f(x)]^2 + O(h^6) \\
&= \frac{5}{4nh} - \frac{1}{n}[\frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}]^2 + O(h^6) \\
&= \frac{5}{4nh} - \frac{1}{n}[\frac{1}{2\pi} e^{-x^2}] + O(h^6) \\
&= \frac{5}{4nh} - \frac{1}{2n\pi} e^{-x^2} + 0(h^6)
\end{aligned}$$