# Problem 1

## a)

Show that the kernel density estimate

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

with kernel $K$ and bandwidth $h > 0$, is a valid density. What condition(s) did you require on $K$?

#### Answer

For $\hat{f}$ to be a valid density, it must be nonnegative (over its support) and integrate to one (for $X$ continuous).

Based on class, we generally want to make assumptions of the kernel, and make minimal assumptions about the true density $f_X(x)$. To that end:

Assume the kernel function, $K : \mathrm{R} \to [0, \infty)$ is measurable with $\int_{-\infty}^{\infty} K(u)\, du = 1$. (Our necessary assumptions.)

It then follows, if $K \geq 0$, then $\hat{f}(x) \geq 0$ for all $x$ ($K$ is non-negative, and we are multiplying it by some scalar, which necessarily must also be a non-negative quantity).

We then must satisfy the second property. To that end we evaluate the integral:

$$\begin{aligned}
\int_{-\infty}^{\infty} \hat{f}(x)\, dx &= \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{x - X_i}{h}\right) dx \\
&= \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{x - X}{h}\right) dx \quad \text{Via X's iid} \\
&= \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} K(u)\, du \quad \text{Via u substitution, where u } = \frac{x - X}{h} \\
&= \frac{1}{n} \sum_{i=1}^{n} 1 \quad \text{Using the property } \int_{-\infty}^{\infty} K(u)\, du = 1 \\
&= \frac{n}{n} \\
&= 1
\end{aligned}$$

Hence $\hat{f}$ is a valid probability density function whenever $K$ itself is a density, needing only assume that the kernel $K$ is a proper density.

## b)

Show that the kernel density estimate

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

with kernel $K$ and bandwidth function $h(x) > 0,\ \forall x$, is *not* a valid density.

**Answer**

As given, define a kernel $K$ and bandwidth function $h(x) > 0,\ \forall x$. These will be the sole assumptions made, otherwise, provided enough assumptions, we could define a valid density.

Let $h : \mathrm{R} \to (0, \infty)$ be a bandwidth function of the point $x$.

We still get the first property of a), namely: $K \geq 0$, then $\hat{f}(x) \geq 0$ for all $x$. The potential culprit then is whether we satisfy the other property (normalization, integrates to 1 over the support). To that end, we note the KDE is then given by:

$$\hat{f}(x) \ = \ \frac{1}{n\, h(x)} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h(x)}\right)$$

Such that:

$$\begin{aligned}
\int_{-\infty}^{\infty} \hat{f}(x)\, dx &= \int_{-\infty}^{\infty} \sum_{i=1}^{n} \frac{1}{nh(x)} K\left(\frac{x - X_i}{h(x)}\right) dx \\
&= \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \frac{1}{h(x)} K\left(\frac{x - X_i}{h(x)}\right) dx \quad \text{As the sum is finite, and some moving of terms} \\
&= \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \frac{1}{h(x)} K\left(\frac{x - X}{h(x)}\right) dx \quad \text{Given iid X, though this isn't important for our purposes}
\end{aligned}$$

The issue then becomes whether:

$$\int_{-\infty}^{\infty} \frac{1}{h(x)} K\left(\frac{x - X_i}{h(x)}\right) dx = 1$$

As given, $h$ depends on $x$, meaning trick used in part a) is not valid, i.e., the transformation $u = (x - X_i)/h(x)$ is no longer linear. Instead, we'd have $u = \frac{x-X}{h(x)}$, and notably:

$$du = \frac{h(x) - (x - X)h'(x)}{h(x)^2}$$

Notably, the above du term involves both $h(x)$ and $h'(x)$, such that $dx$ is **not** just a constant multiple of $du$.

It then follows that, without additional assumptions, there is no guarantee that:

$$\int_{-\infty}^{\infty} \frac{1}{h(x)} K\left(\frac{x - X_i}{h(x)}\right) dx = 1$$

and hence why in general the variable bandwidth kernel density estimator is not a valid density when based solely upon the assumptions given.

# Problem 2

A natural estimator for the $r$th derivative $f^{(r)}(x)$ of $f(x)$ is

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

assuming that $K$ satisfies the necessary differentiability conditions.

## a)

Derive an asymptotic expression for the bias of $\hat{f}^{(r)}(x)$. Also mention the assumptions you made to obtain this result.

**Answer**

To calculate bias, we first need to evaluate the expectation:

$$\begin{aligned}
\mathrm{E}\,\hat{f}^{(r)}(x) =& \frac{1}{h^{r+1}} \int K^{(r)}\left(\frac{x-y}{h}\right) f(y)\,dy \\
=& \frac{1}{h^r} \int K^{(r)}(u)\, f(x-hu)\,du \quad \text{Via } u = (x-y)/h \;\Rightarrow\; y = x - hu,\; dy = -h\,du
\end{aligned}$$

We then use a Taylor's series approximation for $f(x - hu)$ around $x$:

$$f(x-hu) = f(x) - huf'(x) + \tfrac{1}{2}h^2u^2f''(x) + \cdots + \frac{(-hu)^{r+2}}{(r+2)!} f^{(r+2)}(x) + o(h^2)$$

Substituting this into the original equation, we have:

$$\begin{aligned}
\mathrm{Bias}\left[\hat{f}^{(r)}(x)\right] &= \mathrm{E}\;\hat{f}^{(r)}(x) - f^{(r)}(x) \quad \text{Via bias formula} \\
&= \frac{1}{h^r}\left(\int K^{(r)}(u)\left(f(x) - huf'(x) + \tfrac{1}{2}h^2u^2f''(x) + \cdots + \frac{(-hu)^{r+2}}{(r+2)!} f^{(r+2)}(x)o(h^2)\right)\right) - f^{(r)}(x) \quad \text{Substitution} \\
&= \frac{\mu_2}{2} f^{(r+2)}(x)\, h^2 + o(h^2) \quad \text{Identifying leading order term}
\end{aligned}$$

Following the text's standard notation where $\mu_2 = \int u^2 K(u)\,du$.

The assumptions are as follows (and are similar to those used in Chapter 2):

(1): $f$ has $r+2$ continuous derivatives (in a neighborhood of $x$, though we could just say absolutely continuous to make our lives easier).
(2): $K$ has finite second moment (K symmetric would also accomplish the same result, though would be more imposing).
(3): $h \to 0$, $nh \to \infty$.

## b)

Derive an asymptotic expression for the variance of $\hat{f}^{(r)}(x)$. Mention the assumptions you made to obtain this result.

### Answer

$$\begin{aligned}\mathrm{Var}\big[\hat{f}^{(r)}(x)\big] &= \frac{1}{n}\,\mathrm{Var}\left(\frac{1}{h^{r+1}}K^{(r)}\left(\frac{x-X}{h}\right)\right) \quad \text{Via iid assumption}\\ &= \frac{1}{n}\left\{\mathrm{E}\left[\frac{1}{h^{2r+2}}\left(K^{(r)}\left(\tfrac{x-X}{h}\right)\right)^2\right] - \left(\mathrm{E}\left[\tfrac{1}{h^{r+1}}K^{(r)}\left(\tfrac{x-X}{h}\right)\right]\right)^2\right\} \quad \text{Using the definition of variance}\end{aligned}$$

For the leading term, compute the expectation by substitution:

$$\begin{aligned}\mathrm{E}\left[\frac{1}{h^{2r+2}}\left(K^{(r)}\left(\tfrac{x-X}{h}\right)\right)^2\right] &= \frac{1}{h^{2r+2}}\int\left(K^{(r)}\left(\tfrac{x-y}{h}\right)\right)^2 f(y)\,dy\\ &= \frac{1}{h^{2r+1}}\int\left(K^{(r)}(u)\right)^2 f(x-hu)\,du \quad \text{Where } u=\frac{(x-y)}{h}, dy=-hdu\\ &= \frac{f(x)}{h^{2r+1}}\int\left(K^{(r)}(u)\right)^2 du + o\left(\tfrac{1}{h^{2r+1}}\right) \quad \text{as h} \to 0\end{aligned}$$

So, returning back to the variance formula, we have:

$$\mathrm{Var}\big[\hat{f}^{(r)}(x)\big] = \frac{f(x)\,R(K^{(r)})}{nh^{2r+1}} + o\left(\tfrac{1}{nh^{2r+1}}\right) \quad \text{As } \left(\mathrm{E}\left[\tfrac{1}{h^{r+1}}K^{(r)}\left(\tfrac{x-X}{h}\right)\right]\right)^2 \text{ Is } O(1)$$

Where: $R(K^{(r)}) = \int\left(K^{(r)}(u)\right)^2 du$, following a similar convention to that used in the text.

The assumptions are as follows,and are similar to those used in Chapter 2 and part a):

(1): $f$ is continuous (absolutely, or at least in a neighborhood of $x$)

(2): $R(K^{(r)}) < \infty$

(3): $h \to 0$ and $nh^{2r+1} \to \infty$

## c)

Calculate the mean squared error (MSE) of $\hat{f}^{(r)}(x)$.

### Answer

Combining squared bias and variance from parts a) and b):

$$\mathrm{MSE}\big(\hat{f}^{(r)}(x)\big) = \left(\tfrac{\mu_2}{2}f^{(r+2)}(x)h^2\right)^2 + \frac{f(x)\,R(K^{(r)})}{nh^{2r+1}} + o\left(h^4 + \tfrac{1}{nh^{2r+1}}\right)$$

## d)

Calculate the mean integrated squared error (MISE) of $\hat{f}^{(r)}$.

**Answer**

Integrating the MSE from part c) gives us:

$$
\begin{aligned}
\text{MISE}\big(\hat{f}^{(r)}\big) &= \int \text{MSE}\big(\hat{f}^{(r)}(x)\big)\,dx \quad \text{definition} \\
&= \int \left[\left(\tfrac{\mu_2}{2} f^{(r+2)}(x)h^2\right)^2 + \frac{f(x)\,R(K^{(r)})}{nh^{2r+1}} + o\big(h^4 + \tfrac{1}{nh^{2r+1}}\big)\right] dx \quad \text{Substituting known quantities} \\
&= \frac{\mu_2^2}{4}\,h^4 \int \big(f^{(r+2)}(x)\big)^2\,dx \;+\; \frac{R(K^{(r)})}{nh^{2r+1}} \int f(x)\,dx \quad \text{Separating terms} \\
&\qquad + \int o\big(h^4 + \tfrac{1}{nh^{2r+1}}\big)\,dx \quad \text{For spacing purposes, isolating the "o" terms} \\
&= \frac{\mu_2^2}{4}\,h^4 \int \big(f^{(r+2)}(x)\big)^2\,dx \;+\; \frac{R(K^{(r)})}{nh^{2r+1}} \;+\; o\big(h^4 + \tfrac{1}{nh^{2r+1}}\big)
\end{aligned}
$$

This agrees with the AMISE expression for $r = 0$.

## e)

From all your previous results, can you conclude why density derivative estimation is becoming increasingly more difficult for estimating higher order derivatives?

**Answer**

From parts b)–d), the variance term is of leading order $1/(nh^{2r+1})$. Specifically:

$$\text{Var}\big[\hat{f}^{(r)}(x)\big] = \frac{f(x)\,R(K^{(r)})}{nh^{2r+1}} + o\big(\tfrac{1}{nh^{2r+1}}\big)$$

As every little-o is also Big-O we may then say:

$$\text{Var}\big[\hat{f}^{(r)}(x)\big] = O\big(\tfrac{1}{nh^{2r+1}}\big)$$

So, as $r$ increases:

(1): The variance increases (for a fixed $h$).

(2): If we wish to reduce variance, we ultimately do so by trading off with increased bias (bias being of order $O(h^2)$)

(3): So you effectively introduce more bias to get a lower variance for higher-order derivations, i.e., the bias–variance tradeoff becomes “more costly”

## f)

Find an expression for the asymptotically optimal constant bandwidth.

**Answer**

We want to minimize the AMISE expression from part d):

$$\text{AMISE}(h) = \frac{\mu_2^2}{4} h^4 \int \left(f^{(r+2)}(x)\right)^2 dx \; + \; \frac{R(K^{(r)})}{nh^{2r+1}} \; + \; o\big(h^4 + \tfrac{1}{nh^{2r+1}}\big)$$

To find the value of h which minimizes the expression, we differentiate with respect to $h$ and set equal to zero:

$$\frac{d}{dh}\,\text{AMISE}(h) = 4\left(\tfrac{\mu_2^2}{4} \int \left(f^{(r+2)}(x)\right)^2\right) h^3 - \frac{(2r+1)(R(K^{(r)}))}{n} h^{-(2r+2)} = 0$$

Gathering terms, and isolating the h, we have the asymptotically optimal constant bandwidth given by:

$$h^*_{\text{AMISE}} = \left[\frac{(2r+1)\,R(K^{(r)})}{\mu_2^2 \int (f^{(r+2)}(x))^2 dx}\right]^{\frac{1}{2r+5}} n^{-\frac{1}{2r+5}}$$

For $r = 0$, this reduces to the “typical” optimal bandwidth expression given in Chapter 2.