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:

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

This is to say that $\hat{f}$ is a valid probability density function whenever K itself is a density, such that the only necessary assumption(s) are that the kernel K is a proper (valid) 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.

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 (not a linear transformation).

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 (there is dependence on the bandwidth function $h(x)$, which would necessitate additional assumptions to ensure $\hat{f}(x)$ is a valid density).

Note: An alternative approach we could take is to define some bandwidth function that satisfies $h(x) > 0$, $\forall x$, assume $\hat{f}(x)$ is a valid density, and then arrive at some nonsense (for a proof by negation).

To that end, one such function could be $h(x) = |x| + 1$, using a Uniform kernel, with

$$\hat{f}(x) = \frac{1}{2(|x|+1)}$$

This bandwidth function meets our base assumptions, yet:

$$\int_{-\infty}^{\infty} \hat{f}(x)\, dx = \int_{-\infty}^{\infty} \frac{1}{2(|x|+1)}\, dx = \int_{0}^{\infty} \frac{1}{|x|+1}\, dx = \infty$$

So, clearly $\hat{f}(x)$ is not a valid density.

Problem 2

A natural estimator for the rth 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

Start with the expectation of the estimator:

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

Where:

$$u = \frac{x - y}{h}, \quad y = x - hu, \quad dy = -h\, du$$

Our goal is to simplify/evaluate $\int K^{(r)}(u)\, f(x - hu)\, du$. To that end note: Via integration by parts (r-many times), for any sufficiently smooth g (see Assumptions),

$$\int K^{(r)}(u)\, g(u)\, du = (-1)^r \int K(u)\, g^{(r)}(u)\, du$$

With $g(u) = f(x - hu)$, $g^{(r)}(u) = (-h)^r f^{(r)}(x - hu)$.

Such that:

$$\int K^{(r)}(u)\, f(x - hu)\, du = h^r \int K(u)\, f^{(r)}(x - hu)\, du$$

Therefore,

$$\mathrm{E}\, \hat{f}^{(r)}(x) = \frac{1}{h^r} h^r \int K(u)\, f^{(r)}(x - hu)\, du = \int K(u)\, f^{(r)}(x - hu)\, du$$

Now seems a good time for a Taylor Series. To that end, expand $f^{(r)}(x - hu)$ around x:

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

Some Assumptions being made at this step:

- $\int K(u), du = 1$,
- $\int uK(u), du = 0$ (e.g. for symmetric K, to make calculations easier),
- $\mu_2 = \int u^2 K(u), du < \infty$, following the notation used in the text.

Taken together, we have:

$$\mathrm{E}\,\hat{f}^{(r)}(x) = f^{(r)}(x) + \frac{\mu_2}{2}\,h^2 f^{(r+2)}(x) + o(h^2)$$

Then, turning back to the original Bias formula:

$$\begin{aligned}
\mathrm{Bias}\big[\hat{f}^{(r)}(x)\big] &= \mathrm{E}\,\hat{f}^{(r)}(x) - f^{(r)} \\
&= f^{(r)}(x) + \frac{\mu_2}{2}\,h^2 f^{(r+2)}(x) + o(h^2) - f^{(r)}(x) \\
&= \frac{\mu_2}{2}\,f^{(r+2)}(x)\,h^2 + o(h^2)
\end{aligned}$$

(Overall) Assumptions:

(1): f has $r+2$ continuous derivatives in a neighborhood of x (could also say "absolutely continuous", though this is a much stronger assumption)

(2):K is a kernel and a valid density (based on allusions made in-class, K need not be a valid density, but instead satisfy being real-valued and $\int K = 1$)

(3): K is r-times differentiable, with derivatives up to order r continuous and integrable

(4): $h \to 0$.

b)

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

Answer

$$\begin{aligned}
\mathrm{Var}\big[\hat{f}^{(r)}(x)\big] &= \frac{1}{n}\,\mathrm{Var}\bigg(\frac{1}{h^{\,r+1}}\,K^{(r)}\Big(\frac{x-X}{h}\Big)\bigg) \quad \text{under iid X's} \\
&= \frac{1}{n}\bigg\{\mathrm{E}\Big[\frac{1}{h^{\,2r+2}}\Big(K^{(r)}\big(\tfrac{x-X}{h}\big)\Big)^2\Big] - \Big(\mathrm{E}\Big[\tfrac{1}{h^{\,r+1}}K^{(r)}\big(\tfrac{x-X}{h}\big)\Big]\Big)^2\bigg\} \quad \text{variance formula}
\end{aligned}$$

As in part a), we use the change of variables where $u = (x-y)/h$, $dy = -h\,du$:

$$\begin{aligned}
\mathrm{E}\Big[\frac{1}{h^{\,2r+2}}\Big(K^{(r)}\big(\tfrac{x-X}{h}\big)\Big)^2\Big] &= \frac{1}{h^{\,2r+2}}\int \Big(K^{(r)}\big(\tfrac{x-y}{h}\big)\Big)^2 f(y)\,dy \\
&= \frac{1}{h^{\,2r+1}}\int \big(K^{(r)}(u)\big)^2 f(x-hu)\,du \\
&= \frac{1}{h^{\,2r+1}}\Big[\,f(x)\int \big(K^{(r)}(u)\big)^2\,du \;+\; o(1)\Big] \quad h\to 0 \\
&= \frac{f(x)}{h^{\,2r+1}}\,R\big(K^{(r)}\big) \;+\; o\Big(\tfrac{1}{h^{\,2r+1}}\Big) \quad \text{noted below}
\end{aligned}$$

where $R(K^{(r)}) = \int (K^{(r)}(u))^2 du$, following similar notation used in the text.

Note on last line: By continuity of f at x we have $f(x-hu) \to f(x)$ pointwise convergence, and by the dominated convergence theorem:

$$\int (K^{(r)})^2 f(x-hu)\,du = f(x)R(K^{(r)}) + o(1)$$

Note: We evaluated the first term in the variance decomposition. For the second term, from part a), we know that

$$\mathbb{E}\Big[\tfrac{1}{h^{\,r+1}}K^{(r)}\big(\tfrac{x-X}{h}\big)\Big] = f^{(r)}(x) + O(h^2)$$

so its square is $O(1)$ and, after multiplying by $1/n$, contributes $O(1/n)$; and since $h \to 0$, noting little o arithmetic properties:

$$\frac{O(1/n)}{1/(nh^{2r+1})} \;=\; O(h^{2r+1})$$

meaning $O(1/n) = o\big(1/(nh^{2r+1})\big)$. Therefore the squared-mean term is negligible relative to the leading term from the first component of the variance decomposition.

Leaving us with an overall Variance expression of the form:

$$\mathrm{Var}\big[\hat{f}^{(r)}(x)\big] = \frac{f(x)\,R(K^{(r)})}{n\,h^{\,2r+1}} \;+\; o\Big(\tfrac{1}{n\,h^{\,2r+1}}\Big)$$

Assumptions on next page

Assumptions:

(1): f is continuous at x

(2): $R\big(K^{(r)}\big) = \displaystyle\int (K^{(r)}(u))^2 \, du < \infty$

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

c)

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

Answer

Combining squared bias and variance from parts a) and b), and gathering terms for the remainder error term:

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

d)

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

Answer

Integrating the MSE from part c) gives us:

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

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 (not the other way around though!) we may then say:

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

As r increases (and for a fixed h):

(1): The variance increases.

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

(3): So we 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:

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