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 and integrate to one.

- Assume $K : \mathbb{R} \to [0, \infty)$ is measurable with $\int_{-\infty}^{\infty} K(u)\, du = 1$.
- If $K \geq 0$, then clearly $\hat{f}(x) \geq 0$ for all x.

For the integral:

$$\begin{aligned}
\int_{-\infty}^{\infty} \hat{f}(x)\, dx &= \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{x - X_i}{h}\right) dx \\
&\overset{u=(x-X_i)/h}{=} \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} K(u)\, du \\
&= \frac{1}{n} \sum_{i=1}^{n} 1 = 1.
\end{aligned}$$

Hence $\hat{f}$ is a valid probability density function whenever K itself is a 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

Positivity still holds if $K \geq 0$, so $\hat{f}(x) \geq 0$.
The issue lies in normalization:

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

When h depends on x, the substitution $u = (x - X_i)/h(x)$ is not a simple linear map—its Jacobian involves $h(x)$ and $h'(x)$. Consequently, the integral of each term does not generally equal 1. The estimator is therefore not guaranteed to integrate to one.

Counterexample.
Take $K = \phi$ the standard normal pdf, $X_i = 2$, and

$$h(x) = \begin{cases} 1, & x \leq 0, \\ 2, & x > 0. \end{cases}$$

Then

$$\int_{-\infty}^{\infty} \frac{1}{h(x)}\, \phi\left(\frac{x-2}{h(x)}\right)\, dx = \int_{-\infty}^{0} \phi(x-2)\, dx + \int_{0}^{\infty} \tfrac{1}{2}\, \phi\big(\tfrac{x-2}{2}\big)\, dx.$$

Evaluating,

$$= \Phi(-2) + \Phi(1) \approx 0.0228 + 0.8413 = 0.8641 \neq 1.$$

Thus the integral can be strictly less (or greater) than one, so $\hat{f}$ is not a valid density in general.

Note. This form is often called the *balloon estimator*, and it is distinct from the fixed-bandwidth KDE in part (a). Only the constant-bandwidth version is guaranteed to be 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

We have

$$\mathrm{E}\,\hat{f}^{(r)}(x) = \frac{1}{h^{r+1}} \int K^{(r)}\left(\frac{x - y}{h}\right) f(y)\,dy.$$

With $u = (x - y)/h$ so $y = x - hu,\ dy = -h\,du$,

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

Expand $f(x - hu)$ by Taylor expansion around x:

$$f(x - hu) = \sum_{j \ge 0} \frac{(-hu)^j}{j!}\, f^{(j)}(x).$$

Carrying derivatives inside the convolution (or equivalently integrating by parts r times), the leading bias term is

$$\mathrm{bias}\Big[\hat{f}^{(r)}(x)\Big] = \frac{\mu_2(K)}{2}\, f^{(r+2)}(x)\, h^2 + o(h^2),$$

where $\mu_2(K) = \int u^2 K(u)\,du$.

Assumptions: f has $r + 2$ continuous derivatives in a neighborhood of x; K has finite second moment; $h \to 0$, $nh \to \infty$.

b)

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

Answer

By independence of the sample,

$$\mathrm{Var}\big[\hat{f}^{(r)}(x)\big] = \frac{1}{n}\mathrm{Var}\left(\frac{1}{h^{r+1}}K^{(r)}\left(\frac{x-X}{h}\right)\right).$$

As $h \to 0$,

$$\mathrm{Var}\big[\hat{f}^{(r)}(x)\big] \approx \frac{f(x)}{nh^{2r+1}}\int \big(K^{(r)}(u)\big)^2\,du = \frac{f(x)\,R(K^{(r)})}{nh^{2r+1}} + o\left(\frac{1}{nh^{2r+1}}\right),$$

where $R(K^{(r)}) = \int \big(K^{(r)}(u)\big)^2\,du$.

c)

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

Answer

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

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

d)

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

Answer

Integrating the MSE over x gives

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

This parallels the AMISE expression given 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 b)–d), the variance term scales as $1/(nh^{2r+1})$. As r increases:

- The variance grows more quickly for a fixed h.
- To control variance one must increase h, but that worsens the $O(h^2)$ bias.
- Thus the bias–variance tradeoff deteriorates with r.

This explains why estimating higher derivatives is increasingly difficult.

f)

Find an expression for the asymptotically optimal constant bandwidth.

Answer

Let

$$\text{AMISE}(h) = Ah^4 + \frac{B}{nh^{2r+1}},$$

with

$$A = \tfrac{\mu_2(K)^2}{4} \int \left(f^{(r+2)}(x)\right)^2 dx, \qquad B = R(K^{(r)}).$$

Differentiate and set to zero:

$$4Ah^3 - \frac{(2r+1)B}{n} h^{-(2r+2)} = 0 \quad \Rightarrow \quad h^{2r+5} = \frac{(2r+1)B}{4An}.$$

Therefore,

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

For $r = 0$, this reduces to the familiar bandwidth expression from the notes.