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?

We need nonnegativity and unit integral. If K is a probability density on $\mathbb{R}$ (i.e., $K(u) \geq 0$ a.e. and $\int K(u)\,du = 1$), then each term $x \mapsto \frac{1}{h}K\big(\frac{x-X_i}{h}\big)$ is a shifted, rescaled pdf. A uniform mixture of pdfs is a pdf:

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

and $\hat{f}(x) \geq 0$. Hence $\hat{f}$ is a density when K itself is a density.

Answer: Take $K \geq 0$ and $\int K = 1$ (i.e., K a pdf). Then $\hat{f}$ 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.

Check the integral:

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

With h depending on x, the change of variables $u = (x - X_i)/h(x)$ is not a simple linear map (Jacobian is not constant and involves h and h'). Thus $\int \frac{1}{h(x)}K\big(\frac{x-X_i}{h(x)}\big)\,dx$ does not generally equal 1, so the overall integral need not be 1. Therefore $\hat{f}$ may fail to integrate to 1 (and can even be < 1 or > 1). (The notes distinguish local variable bandwidth $h(x)$ from constant h precisely as a different estimator; only the constant-h KDE is guaranteed to be a density under K a pdf.)

Answer: In general $\int \hat{f} \neq 1$ when $h = h(x)$, so this estimator 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.

Write $\mathrm{E}\, \hat{f}^{(r)}(x) = \frac{1}{h^{r+1}} \int K^{(r)}\left(\frac{x-y}{h}\right) f(y)\, dy$. Let $u = (x-y)/h \Rightarrow y = x - hu$, $dy = -h\, du$:

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

Expand $f(x - hu)$ in Taylor series around x: $f(x - hu) = \sum_{j \geq 0} \frac{(-hu)^j}{j!} f^{(j)}(x)$. Integration by parts r times (or using that differentiation w.r.t. x passes inside the convolution) yields the leading terms (paralleling the $r = 0$ case where bias $\sim \frac{1}{2}\mu_2(K) f''(x) h^2$):

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

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

Assumptions: f has $r + 2$ continuous derivatives near x; K has finite second moment and $nh \to \infty$, $h \to 0$. (Pattern follows the lecture's $r = 0$ derivation.)

(b)

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

As in the notes for $r = 0$,

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

so

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

This mirrors the $r = 0$ formula $\mathrm{Var}[\hat{f}(x)] = \frac{f(x) R(K)}{nh} + o((nh)^{-1})$ from the notes.

(c)

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

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

(Compare with the $r = 0$ MSE in the notes.)

(d)

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

Integrate the MSE over x (assuming $f^{(r+2)} \in L^2$):

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

This is the direct analogue of the AMISE expression given for $r = 0$ in the slides.

(e)

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

From (b)–(d), the variance term scales as $1/(nh^{2r+1})$. As r increases, for a given h the variance explodes (the kernel derivatives are rougher and amplify noise). To control variance you must increase h, but that worsens the bias term $O(h^2)$ times $\|f^{(r+2)}\|$. Hence the bias–variance tradeoff deteriorates with r. (The same phenomenon is emphasized in the $r = 0$ discussion of how h balances bias and variance.)

(f)

Find an expression for the asymptotically optimal constant bandwidth.

Minimize the leading-term MISE in (d):

$$\mathrm{AMISE}(h) = A\, h^4 + \frac{B}{n\, h^{2r+1}}, \quad A = \frac{\mu_2(K)^2}{4} \int \big(f^{(r+2)}\big)^2, \quad B = R\big(K^{(r)}\big).$$

Set derivative to zero:

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

Thus

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

For $r = 0$ this reduces to the familiar $h_{AMISE} = \big[\frac{R(K)}{\mu_2(K)^2 R(f'')}\big]^{1/5} n^{-1/5}$ given in the notes.