Problem 1

Consider the kernel density estimator with $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} X$:

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

and denote

$$(f * g)(x) = \int f(x - y) g(y) \, dy.$$

a)

Show that the exact bias of the kernel density estimator is given by

$$\mathrm{E}[\hat{f}_h(x)] - f(x) = (K_h * f)(x) - f(x).$$

Answer

$$\begin{aligned}
\mathrm{E}[\widehat{f}(x)] &= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i)\right] \\
&= \sum_{i=1}^{n} \frac{1}{n} \mathrm{E}\left[K_h(x - X_i)\right] \quad \text{Expectation is a linear function} \\
&= \mathrm{E}\left[K_h(x - X)\right] \quad \text{X's i.i.d., specifically identical} \\
&= \int_{\mathbb{R}} K_h(x - y) f(y) dy \quad \text{See Note} \\
&= (K_h * f)(x) \quad \text{Convolution definition}
\end{aligned}$$

Note: The penultimate step follows from the definition of expectation for a continuous R.V., where if Y has density f, then $\mathrm{E}g(Y) = \int g(y) f(y) \, dy$. Then, as noted we use the given convolution formula.

Returning then to the bias formula, it then follows:

$$\mathrm{E}[\hat{f}_h(x)] - f(x) = (K_h * f)(x) - f(x)$$

b)

Show that the exact variance of the kernel density estimator equals

$$\mathrm{Var}(\hat{f}_h(x)) = \frac{1}{n}\Big[(K_h^2 * f)(x) - (K_h * f)^2(x)\Big].$$

Answer

To make our lives easier, well maybe not you since you're grading this, define the R.V. $Z_i = K_h(x - X_i)$ (for notational convenience).

Then the kernel density estimator is equivalent to $\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i) = \frac{1}{n}\sum_{i=1}^{n} Z_i$.

Notably, as X's are i.i.d., then the Z's are i.i.d., as defined.

Evaluating the (exact) Variance then:

$$\begin{aligned}
\operatorname{Var}(\hat{f}(x)) &= \operatorname{Var}\left(\tfrac{1}{n}\sum_{i=1}^{n} Z_i\right) \\
&= \tfrac{1}{n}\operatorname{Var}(Z_1) \quad \text{(sum of the variance of i.i.d. R.V.'s)} \\
&= \tfrac{1}{n}\left(\mathrm{E}[Z_1^2] - (\mathrm{E}[Z_1])^2\right) \quad \text{Variance definition/decomposition} \\
&= \tfrac{1}{n}\left(\mathrm{E}[K_h^2(x - X_1)] - (\mathrm{E}[K_h(x - X_1)])^2\right) \quad \text{Substituting original definitionb of } Z_i \\
&= \tfrac{1}{n}\left(\int_{\mathbb{R}} K_h^2(x-y)\, f(y)\, dy - \left\{\int_{\mathbb{R}} K_h(x-y)\, f(y)\, dy\right\}^2\right) \quad \text{Convolution definition} \\
&= \tfrac{1}{n}\left[(K_h^2 * f)(x) - (K_h * f)^2(x)\right]
\end{aligned}$$

Notably, the above also uses the definition of expectation for absolutely continuous R.V.'s like in part a).

c)

Calculate the exact mean squared error (MSE) of the kernel density estimator.

Answer

The formula for the MSE is given by:

$$\operatorname{MSE}(\hat{f}(x)) = \operatorname{Var}(\hat{f}(x)) + \operatorname{Bias}^2(\hat{f}(x))$$

Plugging in the results from a) and b) gives us:

$$\operatorname{MSE}(\hat{f}(x)) = \frac{1}{n}\left[(K_h^2 * f)(x) - (K_h * f)^2(x)\right] + \left[(K_h * f)(x) - f(x)\right]^2$$

You *could* simplify this somewhat, which would amount to:

$$\operatorname{MSE}(\hat{f}(x)) = \tfrac{1}{n}(K_h^2 * f)(x) + \left(1 - \tfrac{1}{n}\right)(K_h * f)^2(x) - 2f(x)(K_h * f)(x) + f(x)^2$$

But honestly, that doesn't seem as nice now, does it?

d)

Calculate the exact mean integrated squared error (MISE) of the kernel density estimator.

Answer

$$\text{MISE}(\hat{f}) = \int_{\mathbb{R}} \text{MSE}(\hat{f}(x))dx$$

Using the result from c), i.e., the original, "unsimplified version":

$$\text{MISE}(\hat{f}) = \frac{1}{n}\left[\int_{\mathbb{R}} (K_h^2 * f)(x)\,dx - \int_{\mathbb{R}} (K_h * f)^2(x)\,dx\right] + \int_{\mathbb{R}} \big[(K_h * f)(x) - f(x)\big]^2\,dx$$

Evaluating the first integral of the above:

$$\begin{aligned}
\int_{\mathbb{R}} (K_h^2 * f)(x)\,dx &= \int_{\mathbb{R}}\int_{\mathbb{R}} K_h^2(x-y)\,f(y)\,dy\,dx \\
&= \int_{\mathbb{R}} f(y)\left\{\int_{\mathbb{R}} K_h^2(x-y)\,dx\right\}dy \qquad \text{Fubini to swap order of integration} \\
&= \int_{\mathbb{R}} f(y)\left\{\int_{\mathbb{R}} K_h^2(u)\,du\right\}dy \qquad \text{u substitution where } u = x-y, du = dx \\
&= \left(\int_{\mathbb{R}} f(y)\,dy\right)\left(\int_{\mathbb{R}} K_h^2(u)\,du\right) \\
&= \int_{\mathbb{R}} K_h^2(u)\,du \quad \text{as we integrate f(y) over its support}
\end{aligned}$$

Because we used Fubini we then are assuming that the function is Lebesgue integrable, which is a given when we assume f is a (valid) density.

Note then, that the squared kernel density is of the form:

$$\int_{\mathbb{R}} (K_h^2 * f)(x)\,dx = \int_{\mathbb{R}} K_h^2(u)\,du = \int_{\mathbb{R}} \frac{1}{h^2}\,K^2\left(\frac{u}{h}\right)\,du$$

Consider an additional change of variables, where $v = u/h$, and $du = h\,dv$.

Then:

$$\int_{\mathbb{R}} \frac{1}{h^2}K^2\left(\frac{u}{h}\right)\,du = \int_{\mathbb{R}} \frac{1}{h^2}\left(K^2(v)\,hdv\right) = \frac{1}{h}\int_{\mathbb{R}} K^2(v)\,dv$$

Notably, up until this point the simplification/evaluation was for the first integral of the original MISE expression.

I do not believe the other two integrals evaluate/simplify nicely, and thus will be left to a form of simplification more akin to notational convenience. We then have the overall (exact) MISE is of the form:

$$\text{MISE}(\hat{f}) = \frac{1}{nh}\int_{\mathbb{R}} K^2(u)\,du - \frac{1}{n}\int_{\mathbb{R}} (K_h * f)^2(x)\,dx + \int_{\mathbb{R}} \big[(K_h * f)(x) - f(x)\big]^2\,dx$$

We can simplify this somewhat, following the convention of the text to define $R(K) = \int_{\mathbb{R}} K(x)^2\,dx$:

$$\text{MISE}(\hat{f}) = \frac{1}{nh}\,R(K) - \frac{1}{n}\,R(K_h * f) + R\big((K_h * f) - f\big)$$

Problem 2

a)

Use Hoeffding's inequality to bound the probability that the kernel density estimator $\hat{f}_h$ deviates from its expectation at a fixed point x, i.e., find an upper bound for

$$P\Big(\big|\hat{f}_h(x) - \mathrm{E}[\hat{f}_h(x)]\big| > \epsilon\Big)$$

for some ϵ, and show how the bound depends on n, h, ϵ and $|K|_\infty = \sup_{u \in \mathrm{R}} |K(u)| < \infty$.

Hint: Hoeffding's inequality states that for i.i.d. random variables Y_i such that $a \leq Y_i \leq b$,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} Y_i - \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right]\right| > \epsilon\right) \leq 2\exp\left(-\frac{2n\epsilon^2}{(b-a)^2}\right).$$

Answer

Starting with our typical form of the kernel and kernel density estimator, let:

$$Y_i = K_h(x - X_i) = \frac{1}{h} K\left(\frac{x - X_i}{h}\right) \qquad \text{where } i = 1, \ldots, n,$$

Then, we may write the kernel density estimator as:

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

Since $|K|_\infty = \sup_{u \in \mathbb{R}} |K(u)| < \infty$, we have bounds given by:

$$-\frac{|K|_\infty}{h} \leq Y_i \leq \frac{|K|_\infty}{h}$$

Thus we may take (noting the hint):

$$a = -\frac{|K|_\infty}{h}, \qquad b = \frac{|K|_\infty}{h}, \qquad (b-a)^2 = \frac{4|K|_\infty^2}{h^2}.$$

Applying Hoeffding's inequality:

$$P\left(\big|\hat{f}_h(x) - \mathrm{E}[\hat{f}_h(x)]\big| > \epsilon\right) \leq 2\exp\left(-\frac{2n\epsilon^2}{(b-a)^2}\right)$$

Simplifying the right-hand side of the inequality:

$$2\exp\left(-\frac{2n\epsilon^2}{4|K|_\infty^2/h^2}\right) = 2\exp\left(-\frac{nh^2\epsilon^2}{2|K|_\infty^2}\right)$$

So

$$P\left(\big|\hat{f}_h(x) - \mathrm{E}[\hat{f}_h(x)]\big| > \epsilon\right) \leq 2\exp\left(-\frac{nh^2\epsilon^2}{2|K|_\infty^2}\right)$$

b)

Suppose you want to construct a uniform bound over a compact interval $[a, b]$. Show that

$$P\left(\sup_{x\in[a,b]} \left|\hat{f}_h(x) - \mathrm{E}[\hat{f}_h(x)]\right| > \epsilon\right) \leq \text{something small.}$$

Write down all the assumptions you're making in the process.

Hint: For a given $\delta > 0$, construct a finite set $N_\delta \subset [a, b]$ such that:

- For every $x \in [a, b]$, there exists $x' \in N_\delta$ with $|x - x'| \leq \delta$
- $|N_\delta| \leq \left\lceil \frac{b-a}{\delta} \right\rceil + 1$

Answer

(1): As $n \to \infty$, $h \to 0$ with $nh^2 \to \infty$.

(2): $X_1, \ldots, X_n$ are i.i.d. with density f (ensures we can apply Hoeffding's inequality).

(3): K is bounded, $|K|_\infty = \sup_{u\in\mathbb{R}} |K(u)| < \infty$.

(4): K is Lipshitz continuous (For a somewhat stronger assumption, we could instead say K is differentiable with bounded derivative, $|K'|_\infty = \sup_{u\in\mathbb{R}} |K'(u)| < \infty$).

Note: The stronger version of Condition 4 implies the kernel density estimator (and its expectation) is Lipschitz continuous on the compact set $[a, b]$. This Lipschitz assumption let us reduce from a supremum over a (possibly) infinite set to a maximum over a finite net.

Now, onto the problem:

Define:

$$Y_i(x) = \frac{1}{h} K\left(\frac{x - X_i}{h}\right)$$

Then, by the Mean Value Theorem:

$$|Y_i(x) - Y_i(x')| \leq \frac{|K'|_\infty}{h^2} |x - x'| \quad \Rightarrow \quad |\hat{f}_h(x) - \hat{f}_h(x')| = \left|\frac{1}{n}\sum_{i=1}^{n}\big(Y_i(x) - Y_i(x')\big)\right| \leq \frac{|K'|_\infty}{h^2} |x - x'|$$

Giving us:

$$\left|\hat{f}_h(x) - \hat{f}_h(x')\right| \leq \frac{|K'|_\infty}{h^2} |x - x'|$$

Taking expectations then,

$$|\mathrm{E}\hat{f}_h(x) - \mathrm{E}\hat{f}_h(x')| \leq \frac{|K'|_\infty}{h^2} |x - x'|$$

(Noting the terms on the right-side of the inequality are non-random, i.e., fixed)

Fix $\delta > 0$. Construct a δ-net $N_\delta \subset [a, b]$ so that

$$|N_\delta| \le \left\lceil \frac{b-a}{\delta} \right\rceil + 1, \qquad \forall x \in [a,b],\ \exists x' \in N_\delta :\ |x - x'| \le \delta$$

For any $x \in [a,b]$ and its nearby grid point $x' \in N_\delta$,

$$\left|\hat{f}_h(x) - \mathrm{E}\hat{f}_h(x)\right| \le \left|\hat{f}_h(x) - \hat{f}_h(x')\right| + \left|\hat{f}_h(x') - \mathrm{E}\hat{f}_h(x')\right| + \left|\mathrm{E}\hat{f}_h(x') - \mathrm{E}\hat{f}_h(x)\right| \ \le \frac{2|K'|_\infty}{h^2}\,\delta + \left|\hat{f}_h(x') - \mathrm{E}\hat{f}_h(x')\right|$$

Where the first and last terms are bounded using the Lipschitz condition.

(Note: The additional terms come from "adding zeros" via $\pm\hat{f}_h(x') \pm \mathrm{E}\hat{f}_h(x')$, followed by the Triangle Inequality)

Choose

$$\delta = \frac{\epsilon h^2}{4|K'|_\infty} \quad \Rightarrow \quad \frac{2|K'|_\infty}{h^2}\,\delta = \frac{\epsilon}{2}$$

Then

$$\left\{ \sup_{x \in [a,b]} \left|\hat{f}_h(x) - \mathrm{E}\hat{f}_h(x)\right| > \epsilon \right\} \subseteq \left\{ \max_{x' \in N_\delta} \left|\hat{f}_h(x') - \mathrm{E}\hat{f}_h(x')\right| > \tfrac{\epsilon}{2} \right\}$$

By the union bound,

$$\mathrm{P}\left(\sup_{x \in [a,b]} \left|\hat{f}_h(x) - \mathrm{E}\hat{f}_h(x)\right| > \epsilon \right) \le |N_\delta| \max_{x' \in N_\delta} \mathrm{P}\left(\left|\hat{f}_h(x') - \mathrm{E}\hat{f}_h(x')\right| > \tfrac{\epsilon}{2} \right)$$

From part a), Hoeffding's inequality gives for each x':

$$\mathrm{P}\left(\left|\hat{f}_h(x') - \mathrm{E}\hat{f}_h(x')\right| > \tfrac{\epsilon}{2} \right) \le 2\exp\left(-\frac{nh^2\epsilon^2}{8|K|_\infty^2} \right)$$

$$\mathrm{P}\left(\sup_{x \in [a,b]} \left|\hat{f}_h(x) - \mathrm{E}\hat{f}_h(x)\right| > \epsilon \right) \le \left(\left\lceil \frac{4(b-a)|K'|_\infty}{\epsilon h^2} \right\rceil + 1 \right) \cdot 2\exp\left(-\frac{nh^2\epsilon^2}{8|K|_\infty^2} \right)$$

We then need to determine whether this term is "something small". To that end note that from the bound

$$\mathrm{P}\left(\sup_{x \in [a,b]} \left|\hat{f}_h(x) - \mathrm{E}\hat{f}_h(x)\right| > \epsilon \right) \le \left(\left\lceil \frac{4(b-a)\,|K'|_\infty}{\epsilon h^2} \right\rceil + 1 \right) \cdot 2\exp\left(-\frac{nh^2\epsilon^2}{8\,|K|_\infty^2} \right)$$

Then, for any fixed $\epsilon > 0$,

$$\left\lceil \frac{4(b-a)\,|K'|_\infty}{\epsilon h^2} \right\rceil + 1 \ \le\ \frac{4(b-a)\,|K'|_\infty}{\epsilon h^2} + 1 \le \frac{C_1}{\epsilon h^2}$$

For some constant $C_1 = 4(b-a)|K'|_\infty + 1$

Hence, for $c_1 = \frac{1}{8|K|_\infty^2}$,

$$\mathrm{P}\left(\sup_{x\in[a,b]}\left|\hat{f}_h(x)-\mathrm{E}\hat{f}_h(x)\right|>\epsilon\right) \;\le\; \frac{2C_1}{\epsilon h^2}\,\exp\bigl(-c_1\,nh^2\epsilon^2\bigr)$$

Since $h \equiv h_n$ satisfies $nh^2 \to \infty$

$$\frac{2C_1}{\epsilon h^2}\,\exp\bigl(-c_1\,nh^2\epsilon^2\bigr) \underset{nh^2\to\infty}{\to} 0$$

Such that:

$$\mathrm{P}\left(\sup_{x\in[a,b]}\left|\hat{f}_h(x)-\mathrm{E}\hat{f}_h(x)\right|>\epsilon\right) \underset{nh^2\to\infty}{\to} 0$$

And we have our desired outcome:

$$\mathrm{P}\left(\sup_{x\in[a,b]}\left|\hat{f}_h(x)-\mathrm{E}\hat{f}_h(x)\right|>\epsilon\right) \le \text{something small}$$

c)

From Question b), construct a nonparametric uniform $1-\alpha$ confidence band for $\mathrm{E}[\hat{f}_h(x)]$, i.e., find $L(x)$ and $U(x)$ such that

$$P\bigl(L(x) \le \mathrm{E}[\hat{f}_h(x)] \le U(x),\ \forall x\bigr) \ge 1-\alpha.$$

Answer

For notational convenience, let $\Lambda = |K'|_\infty/h^2$.

Then, from part b), for any $\delta > 0$ and any δ-net $N_\delta \subset [a,b]$,

$$\left\{\sup_{x\in[a,b]}\left|\hat{f}_h(x)-\mathrm{E}\hat{f}_h(x)\right|>\varepsilon\right\} \subseteq \left\{\max_{x'\in N_\delta}\left|\hat{f}_h(x')-\mathrm{E}\hat{f}_h(x')\right|>\varepsilon-2\Lambda\,\delta\right\}$$

Applying Hoeffding's Inequality at each $x' \in N_\delta$ and the union bound, for any $t > 0$,

$$\mathrm{P}\Bigl(\sup_{x\in[a,b]}\left|\hat{f}_h(x)-\mathrm{E}\hat{f}_h(x)\right|>t+2\Lambda\delta\Bigr) \le 2\,|N_\delta|\exp\Bigl(-\tfrac{nh^2t^2}{8\,|K|_\infty^2}\Bigr)$$

Let

$$m_\delta = \left\lceil \tfrac{b-a}{\delta}\right\rceil + 1, \quad \text{and } t_\alpha(\delta) = \sqrt{\frac{8\,|K|_\infty^2}{n\,h^2}\,\log\Bigl(\frac{2\,m_\delta}{\alpha}\Bigr)}$$

Then

$$\mathrm{P}\Bigl(\sup_{x\in[a,b]}\left|\hat{f}_h(x)-\mathrm{E}\hat{f}_h(x)\right|\le t_\alpha(\delta)+2\Lambda\delta\Bigr) \ge 1-\alpha$$

Therefore, we may construct a nonparametric uniform $1-\alpha$ confidence band for $\mathrm{E}[\hat{f}_h(x)]$ a $(1-\alpha)$ (on a compact interval $[a, b]$) via $(L(x), U(x))$, where:

$$L(x) = \hat{f}_h(x) - \big(t_\alpha(\delta) + 2\Lambda\delta\big), \quad U(x) = \hat{f}_h(x) + \big(t_\alpha(\delta) + 2\Lambda\delta\big)$$

(Using Λ and $t_\alpha(\delta)$ as defined previously.)