STAT 5460: Homework III (Technically II)

Sam Olson

Problem 1

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

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

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

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

Answer

$$\begin{aligned}
\mathrm{E}[\widehat{f}(x)] &= \mathrm{E}\left[\frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i)\right] \\
&= \sum_{i=1}^{n} \frac{1}{n} \mathrm{E}\left[K_h(x - X_i)\right] \quad \text{Expectation is a linear function} \\
&= \mathrm{E}\left[K_h(x - X_1)\right] \quad \text{X's iid, specifically identical} \\
&= \int K_h(x - y) f(y) dy \quad \text{See Note} \\
&= (K_h * f)(x) \quad \text{Convolution definition}
\end{aligned}$$

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, we simply call upon the base convolution formula.

Returning then to the bias formula, it then follows:

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

b)

Show that the exact variance of the kernel density estimator equals

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

Answer

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

Then the kernel density estimator is equivalent to $\hat{f}(x) = \frac{1}{n}\sum_{i=1}^n Z_i$. Notably, as X's are iid, then the Z's are iid, as defined.

Evaluating the exact formula for Variance then:

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

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 exact formula for the MSE is given by:

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

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

$$\mathrm{MSE}(\hat{f}(x)) = \frac{1}{n}\Big[(K_h^2 * f)(x) - (K_h * f)^2(x)\Big] + \big[(K_h * f)(x) - f(x)\big]^2$$

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

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

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) (the original, "unsimplified version"):

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

Evaluating the first integral of the above:

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

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

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

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

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

Then:

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

Notably, this simplification/evaluation was for the first integral. I do not believe the other two integrals nicely evaluate, and thus will be left to a form of simplification more akin to notational convenience later on.

Taking the simplifications/evaluations we could muster, the overall MISE expression is of the form:

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

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

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

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 $a \le Y_i \le b$,

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

Answer

Define

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

so that

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

Since $|K| * \infty = \sup * u \in \mathrm{R} |K(u)| < \infty$, we have the almost sure bound

$$-\frac{|K| * \infty}{h} \le Y_i \le \frac{|K| * \infty}{h}$$

Thus in Hoeffding's inequality we can take

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

Applying Hoeffding's inequality gives

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

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

Such that:

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

Dependence: The bound decays exponentially in n and ϵ^2, and is tighter when h is larger (since the summands are bounded by $|K|_\infty/h$).

Remark (if $K \ge 0$): If K is nonnegative, then $0 \le Y_i \le |K| * \infty/h$, so $(b-a)^2 = (|K| * \infty/h)^2$. In that case the exponent improves by a factor of 4:

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

b)

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

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

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 $|x - x'| \le \delta$
- $|N_\delta| \le \lceil \frac{b-a}{\delta} \rceil + 1$

Answer

All Assumptions used in this part:

(1): $(X_1, \ldots, X_n)$ are i.i.d. with some density (used only for independence and expectations).

(2): $(|K| * \infty := \sup_u |K(u)| < \infty)$ (bounded kernel), so $(|Y_i(x)| \le |K| * \infty/h)$ a.s.

(3): (K) is differentiable with $(|K'| * \infty < \infty)$ (Lipschitz smoothness). Then for any (x,x'),

$$|Y_i(x) - Y_i(x')| \le \frac{|K'| * \infty}{h^2}, |x - x'| \quad \Rightarrow \quad |\hat{f}_h(x) - \hat{f}_h(x')| \le \frac{|K'|_\infty}{h^2}, |x - x'|.$$

Onto the problem itself.

To start, define the kernel and kernel density estimator as before:

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

Then,

$$|Y_i(x) - Y_i(x')| \le \frac{|K'| * \infty}{h^2}, |x - x'| \quad \Rightarrow \quad |\hat{f}_h(x) - \hat{f}_h(x')| \le \frac{|K'|_\infty}{h^2}, |x - x'|.$$

Taking expectations yields the same bound for $(|\mathrm{E}\hat{f}_h(x) - \mathrm{E}\hat{f}_h(x')|)$.

Fix ($\delta > 0$) and choose a (δ)-net ($N_\delta \subset [a,b]$) such that $|N_\delta| \le \lceil \frac{b-a}{\delta} \rceil + 1$, and for every $x \in [a,b]$ there exists $x' \in N_\delta$ with $|x - x'| \le \delta$ (thanks, hint!)

It then follows that for any $x \in [a,b]$ and its $x' \in N_\delta$,

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

Choose

$$\delta = \frac{\varepsilon h^2}{4|K'| * \infty} \quad \Rightarrow \quad \frac{2|K'| * \infty}{h^2}$$

And subsequently,

$$\delta = \frac{\varepsilon}{2}$$

Then

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

Apply the union bound and Hoeffding results from part a) at each ($x' \in N_\delta$).

Then, for each fixed x', $|Y_i(x')| \le |K| * \infty/h$ so $(b-a)^2$)in Hoeffding is $(2|K| * \infty/h)^2$.

Hence

$$\Pr\left(\left|\hat{f}_h(x') - \mathrm{E}\hat{f}_h(x')\right| > \tfrac{\varepsilon}{2}\right) \le 2\exp!\left(-\frac{nh^2(\varepsilon/2)^2}{2|K|_\infty^2}\right) = 2\exp!\left(-\frac{nh^2\varepsilon^2}{8|K|_\infty^2}\right).$$

Therefore,

$$\Pr!\left(\sup_{x\in[a,b]} \left|\hat{f}_h(x) - \mathrm{E}\hat{f}_h(x)\right| > \varepsilon\right) \le |N_\delta| \cdot 2\exp!\left(-\frac{nh^2\varepsilon^2}{8|K| * \infty^2}\right) \quad \le \left(\left\lceil \frac{b-a}{\delta} \right\rceil + 1\right), 2\exp!\left(-\frac{nh^2\varepsilon^2}{8|K| * \infty^2}\right).$$

Plugging in ($\delta = \varepsilon h^2/(4|K'| * \infty)$ gives the explicit dependence

$$\Pr!\left(\sup * x \in [a,b] \left|\hat{f}_h(x) - \mathrm{E}\hat{f}_h(x)\right| > \varepsilon\right); \le; \left(\left\lceil \frac{4(b-a)|K'| * \infty}{\varepsilon h^2} \right\rceil + 1\right), 2\exp!\left(-\frac{nh^2\varepsilon^2}{8|K| * \infty^2}\right).$$

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

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

Answer

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

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

where $L = |K'|_\infty / h^2$.

Applying Hoeffding's inequality at each $x' \in N_\delta$ and using the union bound gives

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

For a given $\alpha \in (0,1)$, define

$$m_\delta := \left\lceil \frac{b-a}{\delta} \right\rceil + 1$$

And

$$t_\alpha(\delta) = \sqrt{\frac{8, |K| * \infty^2}{n, h^2} \log!\left(\frac{2m * \delta}{\alpha} \right)}$$

Where $\lceil x \rceil$ is the typical ceiling function.

Then with probability at least $1-\alpha$,

$$\sup_{x\in[a,b]} \left|\hat{f}_h(x) - \mathrm{E}\hat{f}_h(x)\right| \leq t_\alpha(\delta) + 2L\delta$$

Therefore, a uniform $(1-\alpha)$ confidence band for $\mathrm{E}[\hat{f}_h(x)]$, with $x \in [a,b]$ is given by $(L(x), U(x))$ where $L(x)$ and $U(x)$ are defined:

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

And

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