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 $a \leq Y_i \leq b$,

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

Answer

Let

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

so that the kernel density estimator is

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

Since $|K|_\infty = \sup_{u\in\mathbb{R}} |K(u)| < \infty$, we have the almost sure bound

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

Thus we may 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:

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

That is,

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

So

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

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

Special case (nonnegative kernel): If $K \geq 0$, then $0 \leq Y_i \leq |K|_\infty/h$, so $(b-a)^2 = (|K|_\infty/h)^2$, and the exponent improves by a factor of 4:

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

b)

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

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

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

Answer

(1): $X_1, \ldots, X_n$ are i.i.d. with some density on $\mathbb{R}$. (2): The kernel K is bounded: $|K|_\infty := \sup_{u\in\mathbb{R}} |K(u)| < \infty$. (3): The kernel K is differentiable with bounded derivative: $|K'|_\infty = \sup_{u\in\mathbb{R}} |K'(u)| < \infty$. (4): Work on a compact interval $[a,b]$ with a fixed bandwidth $h > 0$.

Given the setup from part a), we know that boundedness gives $|Y_i(x)| \leq |K|_\infty/h$ for all x. We then also know that $|K'|_\infty < \infty$ (assumptions of differentiability), and, by the mean-value theorem:

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

Taking expectations,

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

(Noting the terms on the right-side of the inequality are non-random)

Then, fix some (small) $\delta > 0$, and define a δ-net $N_\delta \subset [a,b]$ by:

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

Then for such x and x',

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

(The additional terms come from adding a "clever zero", and then taking the XXX Inequality)

Choose

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

Hence

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

Applying results (the bound) from part a), for each fixed x' we have

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

Noting that with $|Y_i(x')| \leq |K|_\infty/h$, $(b-a)^2 = (2|K|_\infty/h)^2)$

Then, we have:

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

For some fixed ϵ, since $|N_\delta| \asymp (b-a)|K'|_\infty/(\epsilon h^2)$ and the tail is $2\exp(-c\,nh^2\epsilon^2)$ with $c = 1/(8|K|_\infty^2)$, we have, as $nh^2 \to \infty$,

$$\frac{C}{\epsilon h^2}\,\exp(-c\,nh^2\epsilon^2) \to 0 \quad \text{if}$$

Such that we may say that:

$$\left(\left\lceil \frac{4(b-a)\,|K'|_\infty}{\epsilon h^2} \right\rceil + 1\right)\cdot 2\exp\left(-\frac{nh^2\epsilon^2}{8\,|K|_\infty^2}\right) \equiv \text{something small}$$

And our desired outcome:

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

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

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

Answer

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

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

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

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

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

Let

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

Then

$$P\left(\sup_{x \in [a,b]} \left| \hat{f}_h(x) - \mathrm{E}[\hat{f}_h(x)] \right| \leq t_\alpha(\delta) + 2L\delta \right) \geq 1 - \alpha$$

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

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

And

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

with $L = |K'|_\infty / h^2$ and $t_\alpha(\delta)$ as above.

Notes: You can pick any convenient δ (e.g., $\delta = h$ or $\delta = (b-a)/m$ for some integer m). If you want a slightly tighter (yet simple) band, you can minimize $t_\alpha(\delta) + 2L\delta$ over $\delta > 0$, but that optimization isn't required for validity.

Explanation of Q2

Here's the big picture of what you just did in Problem 2:

What's going on?

We are studying **how far the kernel density estimator (KDE)**

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

deviates from its **expectation** $\mathbb{E}[\hat{f}_h(x)]$.

- The expectation $\mathbb{E}[\hat{f}_h(x)]$ is the **"smoothed truth"** (bias comes from the kernel and bandwidth).
- The deviation $\hat{f}_h(x) - \mathbb{E}[\hat{f}_h(x)]$ measures **random error (variance)**.
- The tools here are **concentration inequalities** (Hoeffding's inequality) and **covering arguments** (nets) to control the maximum deviation over an interval.

Part (a): Pointwise control

- At a **single point** x, write $\hat{f}_h(x)$ as the average of bounded i.i.d. random variables.
- Apply Hoeffding's inequality.
- Conclusion: with high probability, $\hat{f}_h(x)$ is within

  $$O!\left(\frac{1}{h}\sqrt{\frac{1}{n}}\right)$$

  of its mean.
- The probability of a big deviation decays like $\exp(-c, nh^2\varepsilon^2)$.

Part (b): Uniform control over $[a, b]$

- A bound at one point isn't enough — we want **uniform control** across a whole interval.
- Use a **δ-net**: approximate $[a, b]$ by finitely many grid points.
- Because the kernel is Lipschitz (bounded derivative), values at nearby points are close.
- So bounding at grid points suffices to bound everywhere.
- Apply Hoeffding at each grid point, then union bound over the net.
- Result:

  $$P\Big(\sup_{x\in[a,b]} |\hat{f}_h(x) - \mathbb{E}\hat{f}_h(x)| > \varepsilon\Big); \leq; \big(\text{covering factor}\big) \times \exp(-c, nh^2\varepsilon^2).$$

- The exponential dominates the polynomial covering factor, so the probability goes to 0 if $nh^2 \to \infty$.

Part (c): Confidence bands

- From the uniform deviation inequality, we know that with high probability

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

- This lets us build a **uniform confidence band**:

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

- Meaning: with probability at least $1 - \alpha$, the true smoothed expectation $\mathbb{E}[\hat{f}_h(x)]$ lies between $L(x)$ and $U(x)$ simultaneously for all $x \in [a, b]$.

Why we're doing this

- Kernel density estimation is **random**, and we need to know how reliable it is.
- These steps show that:
  1. At each point, $\hat{f}_h(x)$ is tightly concentrated around its expectation.
  2. Uniformly across an interval, the deviations are also controlled, despite infinitely many x.
  3. This leads to **honest confidence bands** for the density curve (or really its expectation), a key tool in nonparametric inference.

You bounded the random error of the KDE using Hoeffding (pointwise), extended it over an interval with covering arguments (uniform control), and then turned that into a practical tool (confidence bands).

Would you like me to also give you a **"one-sentence intuition"** for each part (a), (b), (c), so you can recall the flow without technical detail?