STAT 5460: Homework III FALL 2025

Maximum score is 30 points, due date is Friday, October 3, 2025. Please upload your solutions (CLEAN version) on October 3, 2025 on Canvas.

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

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

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

(a) [2 points] Show that the exact bias of the kernel density estimator is given by

$$\mathbb{E}(\widehat{f}(x)) - f(x) = (K_h * f)(x) - f(x)$$

(b) [5 points] show that the exact variance of the kernel density estimator equals

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

(c) [3 points] Calculate the exact mean squared error (MSE) of the kernel density estimator.

(d) [4 points] Calculate the exact mean integrated squared error (MISE) of the kernel density estimator.

2. (a) [4 points] 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

$$\mathbb{P}\big(\big|\hat{f}_h(x) - \mathbb{E}\,\hat{f}_h(x)\big| > \epsilon\big)$$

for some ϵ and show how the bound depends on n, h, ϵ and $\|K\|_\infty = \sup_{u\in\mathbb{R}} |K(u)| < \infty$. *Hint:* Hoeffding's inequality states that for i.i.d. random variables Y such that $a \leq Y_i \leq b$

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

(b) [8 points] Suppose you want to construct a uniform bound over a compact interval $[a, b]$. Show that

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

Write down all the assumptions you're making in the process. *Hint:* For a given $\delta > 0$, construct a finite set $\mathcal{N}_\delta \subset [a, b]$ such that:

- For every $x \in [a, b]$, there exists $x' \in \mathcal{N}_\delta$ with $|x - x'| \le \delta$
- $|\mathcal{N}_\delta| \le \lceil \frac{b-a}{\delta} \rceil + 1$

(c) [4 points] From Question (b), construct a nonparametric uniform $1 - \alpha$ confidence band for $\mathbb{E}\, \hat{f}_h(x)$, i.e., find $L(x)$ and $U(x)$ such that

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