# MATH524 – Spring 2025

## Problem Set: Week 4

1. (**Kernel Density Estimation with a Fourth Order Kernel**)

   Let $X_1, \ldots, X_n$ be i.i.d. real-valued samples from a distribution with Lebesgue density function $f(x)$. Suppose that $f$ is four times continuously differentiable with $|f^{(r)}(x)| \leq M$ for $0 \leq r \leq 4$.

   Define the fourth order Epanechnikov kernel as $K(u) = \frac{45}{32}\left(1 - \frac{7u^2}{3}\right)(1-u^2)$ for $-1 \leq u \leq 1$. Recall that the kernel density estimator of $f(x)$ is given by

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

   (a) Show that $K$ is indeed a fourth order kernel by proving the following four statements.

   $$\int_{-1}^{1} K(u)\,\mathrm{d}u = 1, \qquad \int_{-1}^{1} uK(u)\,\mathrm{d}u = 0, \qquad \int_{-1}^{1} u^2K(u)\,\mathrm{d}u = 0, \qquad \int_{-1}^{1} u^3K(u)\,\mathrm{d}u = 0.$$

   You may wish to use the fact that $K(u) = K(-u)$.

   (b) Show that the variance is bounded by

   $$\mathbb{V}\left[\widehat{f}(x)\right] \leq \frac{5M}{4nh}.$$

   You may wish to show first that $\int_{-1}^{1} K(u)^2\,\mathrm{d}u = \frac{5}{4}$.

   (c) Show that the bias is bounded by

   $$\left|\mathbb{E}\left[\widehat{f}(x)\right] - f(x)\right| \leq \frac{45Mh^4}{384}.$$

   You may wish to use a third order Taylor expansion with Lagrange remainder for $f$ around the point $x$. Note that $|K(u)| \leq \frac{45}{32}$.

   (d) Use the bias and variance upper bounds to derive an approximate optimal bandwidth which minimizes the mean squared error at each $x$. You may ignore constants which depend on $M$, giving your answer as a function of $n$ only.

   (e) Using this approximate optimal bandwidth, provide an upper bound on the mean squared error of the kernel density estimator as a function of $n$ only.

   (f) Compare the bias, variance, approximate optimal bandwidth and mean squared error upper bound derived above with those arising from an order 2 kernel, ignoring constants. What do you think would happen in the general case where $p \geq 2$ is even? Give any extra regularity assumptions you require.

   (g) Find $u$ which minimizes $K(u)$ and report the associated value of $K(u)$. Propose a modified estimator $\widetilde{f}(x)$ based on $\widehat{f}(x)$ which satisfies:

   i. $\widetilde{f}(x) \geq 0$ for all $x \in \mathbb{R}$ almost surely.

   ii. $\mathbb{E}\left[\left(\widetilde{f}(x) - f(x)\right)^2\right] \leq \mathbb{E}\left[\left(\widehat{f}(x) - f(x)\right)^2\right]$ for all $x \in \mathbb{R}$.

2. (**Influence functions**) Determine the influence function $\varphi(x;\theta)$ for the variance estimator.