

# ECON675: Assignment 2

Anirudh Yadav

October 8, 2018

## Contents

<b>1</b>	<b>Question 1: Kernel Density Estimation</b>	<b>2</b>
1.1	Density derivatives . . . . .	2
1.2	Optimal bandwidth . . . . .	5
1.3	Monte Carlo experiment . . . . .	5
<b>2</b>	<b>Linear smoothers, cross-validation and series</b>	<b>7</b>
2.1	. . . . .	7

# 1 Question 1: Kernel Density Estimation

## 1.1 Density derivatives

I follow the derivation in Hansen's notes. We are interested in estimating

$$f^{(s)}(x) = \frac{d^s}{dx^s} f(x).$$

The natural estimator is

$$\hat{f}^{(s)}(x) = \frac{d^s}{dx^s} \hat{f}(x)$$

Now, we know that  $\hat{f}(x) = \frac{1}{nh} \sum_i K\left(\frac{X_i - x}{h}\right)$ . Thus,

$$\begin{aligned}\hat{f}^{(1)}(x) &= \frac{-1}{nh^2} \sum_{i=1}^n K^{(1)}\left(\frac{X_i - x}{h}\right), \\ \hat{f}^{(2)}(x) &= \frac{1}{nh^3} \sum_{i=1}^n K^{(2)}\left(\frac{X_i - x}{h}\right), \\ &\vdots \\ \hat{f}^{(s)}(x) &= \frac{(-1)^s}{nh^{1+s}} \sum_{i=1}^n K^{(s)}\left(\frac{X_i - x}{h}\right).\end{aligned}$$

Now,

$$\begin{aligned}\mathbb{E}[\hat{f}^{(s)}(x)] &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}\left[\frac{(-1)^s}{h^{1+s}} K^{(s)}\left(\frac{X_i - x}{h}\right)\right] \\ &= \mathbb{E}\left[\frac{(-1)^s}{h^{1+s}} K^{(s)}\left(\frac{X_i - x}{h}\right)\right], \text{ since } X_i \text{ are iid.} \\ &= \int_{-\infty}^{\infty} \frac{(-1)^s}{h^{1+s}} K^{(s)}\left(\frac{z - x}{h}\right) f(z) dz\end{aligned}$$

Next, we want to use integration by parts:  $\int u dv = uv - \int v du$ . Define

$$dv = \frac{(-1)^s}{h^s} \frac{1}{h} K^{(s)}\left(\frac{z - x}{h}\right) \implies v = \frac{(-1)^s}{h^s} K^{(s-1)}\left(\frac{z - x}{h}\right)$$

And

$$u = f(z) \implies du = f^{(1)}(z) dz.$$

Thus,

$$\begin{aligned}\mathbb{E}[\hat{f}^{(s)}(x)] &= \left[ \frac{(-1)^s}{h^s} K^{(s-1)}\left(\frac{z - x}{h}\right) f^{(1)}(z) \right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{(-1)^s}{h^s} K^{(s-1)}\left(\frac{z - x}{h}\right) f^{(1)}(z) dz. \\ &= - \int_{-\infty}^{\infty} \frac{(-1)^s}{h^s} K^{(s-1)}\left(\frac{z - x}{h}\right) f^{(1)}(z) dz\end{aligned}$$

Repeating this  $s$  times give

$$\begin{aligned}\mathbb{E}[\hat{f}^{(s)}(x)] &= (-1)^s \int_{-\infty}^{\infty} \frac{(-1)^s}{h} K\left(\frac{z-x}{h}\right) f^{(s)}(z) dz \\ &= \int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{z-x}{h}\right) f^{(s)}(z) dz\end{aligned}$$

Next, use the following change of variables:  $u = \frac{z-x}{h}$ , which implies  $z = x + hu \implies dz = hdu$ . Thus,

$$\mathbb{E}[\hat{f}^{(s)}(x)] = \int_{-\infty}^{\infty} K(u) f^{(s)}(x + hu) du \quad (1)$$

The next step is to take a Taylor expansion of  $f^{(s)}(x + hu)$  around  $x + hu = x$ , which is valid if  $h \rightarrow 0$ . We get

$$f^{(s)}(x + hu) = f^{(s)}(x) + f^{(s+1)}(x)hu + \frac{1}{2}f^{(s+2)}(x)h^2u^2 + \dots + \frac{1}{P!}f^{(s+P)}(x)h^Pu^P + o(h^P).$$

Substituting this expression back into (1), integrating over each term, and using the fact that  $\int_{-\infty}^{\infty} K(u)du = 1$  and the notation

$$\mu_\ell(K) = \int_{-\infty}^{\infty} u^\ell K(u) du$$

gives

$$\mathbb{E}[\hat{f}^{(s)}(x)] = f^{(s)}(x) + f^{(s+1)}(x)h\mu_1(K) + \frac{1}{2}f^{(s+2)}(x)h^2\mu_2(K) + \dots + \frac{1}{P!}f^{(s+P)}(x)h^P\mu_P(K) + o(h^P).$$

Finally, noting that since  $K$  is a  $P$ -order kernel,  $\mu_\ell(K) = 0$  for all  $\ell < P$ , gives the desired result

$$\mathbb{E}[\hat{f}^{(s)}(x)] = f^{(s)}(x) + \frac{1}{P!}f^{(s+P)}(x)h^P\mu_P(K) + o(h^P). \quad (2)$$

Next we consider the variance of the derivative estimator.

$$\begin{aligned}\mathbb{V}[\hat{f}^{(s)}(x)] &= \mathbb{V}\left[\frac{(-1)^s}{nh^{1+s}} \sum_{i=1}^n K^{(s)}\left(\frac{X_i - x}{h}\right)\right] \\ &= \frac{1}{nh^{2+2s}} \mathbb{V}\left[K^{(s)}\left(\frac{X_i - x}{h}\right)\right],\end{aligned}$$

since  $\{X_i\}$  are iid there are no covariance terms and each term has the same variance. Continuing,

$$\begin{aligned}\mathbb{V}[\hat{f}^{(s)}(x)] &= \frac{1}{nh^{2+2s}} \left\{ \mathbb{E}\left[K^{(s)}\left(\frac{X_i - x}{h}\right)^2\right] - \mathbb{E}\left[K^{(s)}\left(\frac{X_i - x}{h}\right)\right]^2 \right\} \\ &= \frac{1}{nh^{2+2s}} \mathbb{E}\left[K^{(s)}\left(\frac{X_i - x}{h}\right)^2\right] - \frac{1}{n} \mathbb{E}\left[\frac{1}{h^{1+s}} K^{(s)}\left(\frac{X_i - x}{h}\right)\right]^2\end{aligned} \quad (3)$$

Now, from above we know that

$$\begin{aligned}\mathbb{E}\left[\frac{1}{h^{1+s}}K^{(s)}\left(\frac{X_i - x}{h}\right)\right] &= f^{(s)}(x) + \frac{1}{P!}f^{(s+P)}(x)h^P\mu_P(K) + o(h^P) \\ &= f^{(s)}(x) + o(1)\end{aligned}$$

since the remainder goes to zero as  $h \rightarrow 0$ . Thus, the second term in (3) is  $O(\frac{1}{n})$ ; i.e. the same order as  $1/n$ . Furthermore  $O(\frac{1}{n})$  is of smaller order than  $O(\frac{1}{nh^{1+2s}})$  since  $h \rightarrow 0$  and  $n \rightarrow \infty$ . Accordingly, we can write

$$\mathbb{V}[\hat{f}^{(s)}(x)] = \frac{1}{nh^{2+2s}}\mathbb{E}\left[K^{(s)}\left(\frac{X_i - x}{h}\right)^2\right] + o\left(\frac{1}{nh^{1+2s}}\right),$$

Thus,

$$\mathbb{V}[\hat{f}^{(s)}(x)] = \frac{1}{nh^{1+2s}}\int_{-\infty}^{\infty}\frac{1}{h}K^{(s)}\left(\frac{z - x}{h}\right)^2 f(z)dz + o\left(\frac{1}{nh^{1+2s}}\right)$$

Again we use the change of variables  $u = \frac{z-x}{h}$  so that

$$\mathbb{V}[\hat{f}^{(s)}(x)] = \frac{1}{nh^{1+2s}}\int_{-\infty}^{\infty}K^{(s)}(u)^2 f(x + hu)du + o\left(\frac{1}{nh^{1+2s}}\right)$$

With the usual Taylor expansion of  $f(x + hu)$  we can write

$$\begin{aligned}\mathbb{V}[\hat{f}^{(s)}(x)] &= \frac{1}{nh^{1+2s}}\int_{-\infty}^{\infty}K^{(s)}(u)^2(f(x) + O(h))du + o\left(\frac{1}{nh^{1+2s}}\right) \\ &= \frac{f(x)}{nh^{1+2s}}\int_{-\infty}^{\infty}K^{(s)}(u)^2du + o\left(\frac{1}{nh^{1+2s}}\right) \\ &= \frac{1}{nh^{1+2s}}f(x)\vartheta_s(K) + o\left(\frac{1}{nh^{1+2s}}\right),\end{aligned}$$

where  $\vartheta_s(K) = \int_{-\infty}^{\infty}K^{(s)}(u)^2du$  as required.

## 1.2 Optimal bandwidth

We have

$$\begin{aligned}\text{AIMSE}[h] &= \int_{-\infty}^{\infty} \left[ \left( h^P \mu_P(K) \cdot \frac{f^{(P+s)}(x)}{P!} \right)^2 + \frac{1}{nh^{1+2s}} \vartheta_s(K) f(x) \right] dx \\ &= h^{2P} \left( \frac{\mu_P(K)}{P!} \right)^2 \vartheta_{s+P}(f) + \frac{1}{nh^{1+2s}} \vartheta_s(K),\end{aligned}$$

since  $f(x)$  integrates to 1 and where  $\vartheta_{s+P}(f) = \int (f^{(P+s)}(x))^2 dx$ . Thus,

$$\begin{aligned}\frac{d}{dh} \text{AIMSE}[h] &= 2Ph^{2P-1} \left( \frac{\mu_P(K)}{P!} \right)^2 \vartheta_{s+P}(f) - (1+2s) \frac{1}{nh^{2+2s}} \vartheta_s(K) = 0 \\ \implies 2Ph^{1+2P+2s} \left( \frac{\mu_P(K)}{P!} \right)^2 \vartheta_{s+P}(f) &= (1+2s) \frac{1}{n} \vartheta_s(K),\end{aligned}$$

which gives the optimal bandwidth

$$h^* = \left[ \frac{1+2s}{2Pn} \left( \frac{P!}{\mu_P(K)} \right)^2 \frac{\vartheta_s(K)}{\vartheta_{s+P}(f)} \right]^{\frac{1}{1+2P+2s}}.$$

A fully data-driven method for estimating  $h^*$  is cross-validation. This procedure attempts to directly estimate the mean-squared error, and then choose the bandwidth which minimizes this estimate. From the lecture notes the cross-validation bandwidth is the value  $h$  which minimizes the criteria

$$\hat{h}_{CV} = \arg \min_h CV(h) = \frac{1}{n^2 h} \sum_{i=1}^n \sum_{j=1}^n (K * K) \left( \frac{X_i - X_j}{h} \right) - \frac{2}{n} \sum_{i=1}^n \hat{f}_{(i)}(X_i)$$

where  $\hat{f}_{(i)}(x_i)$  is the density estimate computed without observation  $X_i$ .

## 1.3 Monte Carlo experiment

(a) First, we want to compute the theoretically optimal bandwidth for  $s = 0$ ,  $n = 1000$ , using the Epanechnikov kernel ( $P = 2$ ), with the following Gaussian DGP:

$$x_i \sim 0.5\mathcal{N}(-1.5, -1.5) + 0.5\mathcal{N}(1, 1)$$

From Table 1 in Hansen's notes,  $\mu_2(K) = 1/5$  and  $\vartheta(K) = 3/5$  for the Epanechnikov kernel. Thus, the only other ingredient we need is  $\vartheta_2(f) = \int [f^{(2)}(x)]^2 dx$  for the above DGP. Note that the second derivative of the normal density with mean  $\mu$  and variance  $\sigma^2$  is

$$\phi_{\mu, \sigma^2}^{(2)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{\sigma^2}\right) \left[ \left(\frac{x-\mu}{\sigma^2}\right)^2 - \frac{1}{\sigma^2} \right]$$

Since differentiation is a linear operation, we have

$$\vartheta_2(f) = \int_{-\infty}^{\infty} [0.5 \times \phi_{-1.5,1.5}^{(2)}(x) + 0.5 \times \phi_{1,1}^{(2)}(x)]^2 dx \approx 0.0388.$$

Finally, we get the theoretically optimal bandwidth

$$h^* = \left[ \frac{1}{2 \times 2 \times 1000} \left( \frac{2!}{1/5} \right)^2 \frac{3/5}{\vartheta_2(f)} \right]^{\frac{1}{1+2 \times 2}} \approx 0.827.$$

## 2 Linear smoothers, cross-validation and series

### 2.1

We are interested in estimating the regression function  $e(x) = \mathbb{E}[y_i|x_i = x]$ . The idea of local polynomial regression is to approximate  $e(x)$  locally by a polynomial of degree  $p$ , and estimate this local approximation by weighted least squares. For each  $x$  we solve

$$\hat{\beta}(x) = \arg \min_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^n [y_i - \beta_0 - \beta_1(x_i - x) - \beta_2(x_i - x)^2 - \dots - \beta_p(x_i - x)^p]^2 K\left(\frac{x_i - x}{h}\right).$$

where

$$\hat{e}(x) = \hat{\beta}_0$$

Note that this is motivated by a Taylor expansion of the true regression function  $e(x_i)$  around  $x$ . And note that the kernel is just a ‘smooth’ way of weighting observations that are close to the evaluation point  $x$ .

More compactly, we write

$$\hat{\beta}(x) = \arg \min_{\beta \in \mathbb{R}^{p+1}} \sum_{i=1}^n [y_i - \mathbf{r}_p(x_i - x)' \beta]^2 K\left(\frac{x_i - x}{h}\right)$$

where  $\mathbf{r}_p(u)$  is an operator such that  $\mathbf{r}_p(u) = (1, u, u^2, \dots, u^p)'$ .