

Econ 675 Assignment 1

Nathan Mather

October 9, 2018

Contents

1	Kernal Density Estimation	1
1.1	Part 1	1
1.2	part 2	2
1.3	Monte Carlo experiment	3

1 Kernal Density Estimation

1.1 Part 1

Start by noting that

$$\hat{f}^{(s)}(x) = \frac{(-1)^s}{nh^{1+s}} \sum_{i=1}^n k^{(s)}\left(\frac{x_i - x}{h}\right)$$

Now taking the expectation of $\hat{f}^{(s)}(x)$ that we can apply the linearity of expectations to move the expectation inside the sum. Then we can use the i.i.d. assumption to show the sum is just n times the expectation. This leaves us with

$$E[\hat{f}^{(s)}(x)] = E\left[\frac{(-1)^s}{h^{1+s}} k^{(s)}\left(\frac{x_i - x}{h}\right)\right] = \int_{-\infty}^{\infty} \frac{(-1)^s}{h^{1+s}} k^{(s)}\left(\frac{z - x}{h}\right) f(z) dz$$

Where the second equality is just by the definition of the expectation. Next we use integration by parts. Note that

$$\int_{-\infty}^{\infty} \frac{(-1)^s}{h^{1+s}} k^{(s)}\left(\frac{z - x}{h}\right) f(z) dz = - \int_{-\infty}^{\infty} \frac{(-1)^s}{h^s} k^{(s-1)}\left(\frac{z - x}{h}\right) f^{(1)}(z) dz$$

Iterating this s times gives us

$$\int_{-\infty}^{\infty} \frac{(-1)^s}{h^{1+s}} k^{(s)}\left(\frac{z - x}{h}\right) f(z) dz = (-1)^s \int_{-\infty}^{\infty} \frac{(-1)^s}{h} k\left(\frac{z - x}{h}\right) f^{(s)}(z) dz = \int_{-\infty}^{\infty} \frac{1}{h} k\left(\frac{z - x}{h}\right) f^{(s)}(z) dz$$

Next we apply change of variables. let $u = \frac{z-x}{h}$ Note that $du = \frac{1}{h} dz$ so we get

$$\int_{-\infty}^{\infty} k(u) f^{(s)}(x + hu) du$$

Next we Taylor expand $f^{(s)}(x + hu)$ to the P^{th} order about x . Recall from properties of the kernel estimator that $\int_{-\infty}^{\infty} k(u)du = 1$ and that $\int_{-\infty}^{\infty} k(u)u^j du = 0$ for all $j \neq p$. This gives us

$$f^{(s)}(x) + \frac{1}{P!} f^{(s+P)}(x) h^P \int_{-\infty}^{\infty} k(u) u^p du + o(h^P) = f^{(s)}(x) + \frac{1}{P!} f^{(s+P)}(x) h^p \mu_P(k) + o(h^P)$$

which is the desired result.

Now for the variance recall again that

$$\hat{f}^{(s)}(x) = \frac{(-1)^s}{nh^{1+s}} \sum_{i=1}^n k^{(s)} \left(\frac{x_i - x}{h} \right)$$

So by the i.i.d. assumption we can get that

$$V \left(\hat{f}^{(s)}(x) \right) = \frac{1}{nh^{2+2s}} V \left(k^{(s)} \left(\frac{x_i - x}{h} \right) \right)$$

$$V \left(\hat{f}^{(s)}(x) \right) = \frac{1}{nh^{2+2s}} V \left(k^{(s)} \left(\frac{x_i - x}{h} \right) \right) \quad (1)$$

$$= \frac{1}{nh^{2+2s}} E \left[\left(k^{(s)} \left(\frac{x_i - x}{h} \right) \right)^2 \right] - \frac{1}{nh^{2+2s}} E \left[\left(k^{(s)} \left(\frac{x_i - x}{h} \right) \right)^2 \right]^2 \quad (2)$$

$$= \frac{1}{nh^{2+2s}} E \left[\left(k^{(s)} \left(\frac{x_i - x}{h} \right) \right)^2 \right] - \frac{1}{n} \left(\frac{1}{h^{1+s}} E \left[\left(k^{(s)} \left(\frac{x_i - x}{h} \right) \right)^2 \right] \right)^2 \quad (3)$$

$$= \frac{1}{nh^{2+2s}} \int_{-\infty}^{\infty} k^{(s)} \left(\frac{x_i - x}{h} \right)^2 f(z) dz + \frac{1}{nh^{2+2s}} f^{(n)}(X)^2 \quad (4)$$

$$= \frac{1}{nh^{1+2s}} \int_{-\infty}^{\infty} k^{(s)}(u)^2 f(x + hu) du + o \left(\frac{1}{nh^{2+2s}} \right) \quad (5)$$

$$= \frac{1}{nh^{1+2s}} \cdot \vartheta_s(K) + o \left(\frac{1}{nh^{2+2s}} \right) \quad (6)$$

1.2 part 2

We start with the following AMISE

$$AIMSE[h] = \int \left[\left(h_n^P \cdot \mu_P(K) \cdot \frac{f^{(P+s)}(x)}{P!} \right)^2 + \frac{1}{nh_n^{1+2s}} \cdot \vartheta_s(K) \cdot f(x) \right] dx$$

Using the ϑ notation so $\vartheta_{P+s}(f) = \int (f^{(P+s)}(x))^2$ and recalling that $f(x)$ integrates to 1 we can rewrite this as

$$= h_n^{2P} \left(\frac{\mu_P(K)}{P!} \right)^2 \vartheta_{P+s}(f) + \frac{\vartheta_s(K)}{nh_n^{1+2s}}$$

Now taking first order conditions and solving for h

$$\frac{d}{dh} AIMSE[h] = 2Ph_n^{2p-1} \left(\frac{\mu_P(K)}{P!} \right)^2 \vartheta_{P+s}(f) - (1 + 2s) \frac{\vartheta_s(K)}{nh_n^{2+2s}} = 0$$

$$\implies 2Ph^{1+2P+2s} \left(\frac{\mu_P(K)}{P!} \right)^2 \vartheta_{P+s}(f) = (1+2s) \frac{\vartheta_s(K)}{n}$$

Thus, we get the AIMSE-optimal bandwidth choice.

$$h_{AIMSE_s} = \left[\frac{(2s+1)(P!)^2}{2P} \frac{\vartheta_s(K)}{\vartheta_{s+P}(f) \cdot \mu_P(K)^2} \frac{1}{n} \right]^{\frac{1}{1+2P+2s}}$$

Least squares cross-validation is a fully automatic data-driven method of selecting the smoothing parameter h . This is based on the principle of selecting bandwidth that minimizes the integrated squared error of the resulting estimate. The estimate used is

$$\hat{h}_{CV} = \arg \min_h \frac{1}{n^2 h} \sum_{i=1}^n \sum_{j=1}^n \bar{k} \left(\frac{X_i - X_j}{h} \right) - \frac{2}{n(n-1)h} \sum_{i=1}^n \sum_{j=1, i \neq j}^n k \left(\frac{X_i - X_j}{h} \right)$$

1.3 Monte Carlo experiment

1.3.1 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:

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

Filling in what we know so far we have :

$$h_{AIMSE_s} = \left[\frac{2 * \vartheta_0(K)}{\vartheta_2(f) \cdot \mu_2(K)^2} \frac{1}{1000} \right]^{\frac{1}{5}}$$

So we need the second moment of K and the first moment of the second derivative of k squared. We can get two of these values from the table in Bruce Hanson's nonparametric notes. Giving us.

$$h_{AIMSE_s} = \left[\frac{\frac{3}{5} \cdot 2}{\vartheta_2(f) \cdot \frac{1}{5}} \frac{1}{1000} \right]^{\frac{1}{5}}$$

The second derivative of the normal density φ with mean μ variance σ^2 is

$$\varphi''_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \left[\left(\frac{(x-\mu)}{\sigma^2} \right)^2 - \frac{1}{\sigma^2} \right]$$

now using the linearity of integrals we can find $\vartheta_2(f)$

$$\vartheta_2(f) = \int_{-\infty}^{\infty} [0.5\varphi''_{1,1}(x) + 0.5\varphi''_{-1.5,1.5}(x)]^2 dx \approx 0.03883397$$

Where the approximation comes from R

Finally, plugging this in gives the theoretically optimal bandwidth is:

$$h^* = 0.8267532$$

1.3.2 b

imse_li	imse_lo	h
0.000166	0.000166	0.5
0.000144	0.000145	0.6
0.000124	0.000126	0.7
0.000119	0.000122	0.8
0.000106	0.000109	0.9
0.000124	0.000127	1
0.00013	0.000133	1.1
0.000151	0.000154	1.2
0.00019	0.000194	1.3
0.000215	0.000218	1.4
0.000263	0.000267	1.5