

# Econ 589 Econometrics: Problem set 2

Ste Rose

November 9, 2019

## 1 Question One.

### 1.1 Part (a).

We consider a nonparametric kernel density estimator with standard normal kernel and fixed  $n$ . The bandwidth  $h$  is effectively a measure of how close we want random points  $x_i$  to be to the point for which we're estimating the density,  $x$  (of course in a more sophisticated way via the use of the kernel; rather than a simple binary choice of close enough to count as the same, the kernel with the bandwidth determines a weight).

If the bandwidth  $h$  tends to infinity, then every point in the support is considered to be “close enough” to count as  $x$ , and in fact the argument of every entry in the sum tends to 0. Since  $k$  is continuous, the numerator in the estimator tends to  $k(0)$ , but the denominator  $h$  tends to infinity, so that the estimator tends to 0.

If the bandwidth  $h$  tends to 0, then every point in the support other than  $x$  (if it occurs in the random sample) is considered to be too far away to count at all as a point close to  $x$ , in which case both the numerator and the denominator of the estimator tend to 0. Intuitively, we expect that as the kernel is a transformation of an exponential, that the numerator tends to 0 faster than the denominator, so that the limit is 0. For all  $x_i \neq x$ , we can use the Taylor expansion of  $\phi$  the normal pdf to find the limit

$$\lim_{h \rightarrow 0} \frac{1}{h} k\left(\frac{x - x_i}{h}\right) = \frac{1}{\sqrt{2\pi}} \lim_{h \rightarrow 0} \exp\left(-\frac{(x - x_i)^2}{h^2}\right) \frac{1}{h}. \quad (1)$$

By definition of the exponential function,

$$e^{\frac{(x - x_i)^2}{h^2}} = 1 + \left(\frac{x - x_i}{h}\right)^2 + O(h^{-4}), \quad (2)$$

where the remainder term here is strictly positive, so that

$$e^{\frac{(x - x_i)^2}{h^2}} > \left(\frac{x - x_i}{h}\right)^2, \quad (3)$$

and hence

$$\frac{h^2}{(x - x_i)^2} > e^{-\left(\frac{x - x_i}{h}\right)^2} > 0. \quad (4)$$

Clearly the limit of the left hand side divided by  $h$  tends to 0, and hence our limit of interest tends to 0. In the case where the random sample does not contain a  $j$  such that  $x_j = x$ , this implies that, given  $n$  is fixed, the estimator tends to 0.

If one is unfortunate enough to have such an  $x_j = x$ , then all terms in the sum tend to 0 except the term corresponding to  $x_j$ , which is

$$\frac{k(0)}{h}, \quad (5)$$

which tends to infinity.

## 1.2 Part (b).

We now consider the same question for the regression density estimator, which is

$$\hat{m}(x) = \frac{\hat{a}(x)}{\hat{f}(x)}, \quad (6)$$

where  $\hat{f}(x)$  is our kernel density estimator and

$$\hat{a}(x) = \frac{1}{nh} \sum_{i \leq n} k\left(\frac{x - x_i}{h}\right) y_i. \quad (7)$$

This implies that

$$\hat{m}(x) = \frac{\sum_{i \leq n} k\left(\frac{x - x_i}{h}\right) y_i}{\sum_{i \leq n} k\left(\frac{x - x_i}{h}\right)}. \quad (8)$$

Annoyingly, it is clear that both  $\hat{a}(x)$  and  $\hat{f}(x)$  will tend to 0 as  $h$  tends to infinity, and 0 as  $h$  tends to zero if  $x$  is not in the sample, and infinity if  $x$  is in the sample. However, given our expression for the regression estimator, we can write

$$\hat{m}(x) = \sum_{i \leq n} y_i \frac{k\left(\frac{x - x_i}{h}\right)}{\sum_{j \leq n} k\left(\frac{x - x_j}{h}\right)}. \quad (9)$$

As  $h$  goes to infinity, each of these terms tend to

$$y_i \frac{k(0)}{nk(0)}, \quad (10)$$

so that the estimator tends to

$$\frac{1}{n} \sum_{i \leq n} y_i, \quad (11)$$

which is the sample mean of  $y_i$ .

We now consider  $h \rightarrow 0$ . The reciprocal of the factor multiplying  $y_i$  is

$$\sum_{j \leq n} \frac{k\left(\frac{x-x_j}{h}\right)}{k\left(\frac{x-x_i}{h}\right)} = 1 + \sum_{j \neq i} \frac{k\left(\frac{x-x_j}{h}\right)}{k\left(\frac{x-x_i}{h}\right)}. \quad (12)$$

We consider one of the terms in this sum,

$$\frac{k\left(\frac{x-x_j}{h}\right)}{k\left(\frac{x-x_i}{h}\right)} = \exp\left(\frac{1}{2h^2}(x_j - x_i)(2x - x_i - x_j)\right). \quad (13)$$

If the coefficient of  $h^{-2}$  is positive for any  $j$ , then the expression in equation (12) tends to infinity. If this is the case for each  $x_i$ , the terms multiplying  $y_i$  in our regression estimator all tend to 0, and hence  $\hat{m}(x)$  tends to 0.

The coefficient is positive if

$$x_i < x_j < 2x - x_i, \quad (14)$$

or

$$2x - x_i < x_j < x_i. \quad (15)$$

If this isn't true for any  $x_j$ , then the expression in equation (12) tends to 1, in which case  $y_i$  is not killed by the limit, and we have

$$\hat{m}(x) \rightarrow \frac{1}{n} \sum_{i \leq n} y_i \mathbb{1}\{\forall j : x_j < x_i \iff 2x \geq x_j + x_i\}. \quad (16)$$

## 2 Question Two.

## 3 Question Three.

## 4 Question Four.

## 5 Question Five.

## 6 Question Six.

## 7 Question Seven.

### 7.1 Part (a).

We consider an RDD estimator using standard kernel regression estimation, so that the parameter is an estimate of the difference of the regression estimators either side of the treatment cutoff. Since the Nadaraya-Watson estimator is upward biased if the conditional expectation  $m$  is upward sloping at the boundary [3], we have that the difference would be exaggerated, and hence the RDD estimate is upward biased.

## 7.2 Part (b).

Beyond new kernel's and new estimators, the only change we can make is to the data, and hence we consider transformations of the data that will reduce the boundary bias.

Our intuition as to why the boundary bias occurs for the kernel distribution estimator is that the estimator “expects” there to be data to both sides of the boundary point, and finding only half as many points as it expects, and hence introduces bias. One method to deal with this could be to add points to the other side of the estimator to make up for this, with the most intuitive choice of points being a reflection of the existing data. In fact, such an approach is taken by Hall and Wehrly (1991), who find using a simulation study that an estimator that reflects data across boundaries and runs as a standard kernel estimator outperforms common boundary-corrected kernels, and performs approximately as well as estimators proposed by Gasser-Muller [1].

## 8 Question Eight.

## 9 Question Nine.

We are asked to consider the advantages of the median absolute deviation as a measure of the success of an estimator in simulation studies over the root mean squared error. The definition of the median absolute deviation (hereafter MAD) is defined[2] as

$$\text{MAD} = \text{median}(x_i - \text{median}(x_i)). \quad (17)$$

That is, it is the median of the distribution of the data minus the median of the data.

## 10 Question Ten.

## References

- [1] P. HALL AND T. E. WEHRLY, *A geometrical method for removing edge effects from kernel-type nonparametric regression estimators*, Journal of the American Statistical Association, 86 (1991), pp. 665–672.
- [2] C. LEYS, C. LEY, O. KLEIN, P. BERNARD, AND L. LICATA, *Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median*, Journal of Experimental Social Psychology, 49 (2013), pp. 764–766.
- [3] J. RACINE, *Bias-corrected kernel regression*, Journal of Quantitative Economics, 17 (2001), pp. 25–42.