Economics 673: Applied Nonparametric Econometrics

Department of Economics, Finance and Legal Studies

University of Alabama

Spring 2024

Midterm — Key

The exam consists of three questions on three pages. Each question is of equal value.

1. We are interested in the optimal bandwidth which minimizes the asymptotic mean integrated square

$$AMISE\left[\widehat{f}(x)\right] = \frac{h^4\kappa_2^2(k)}{4}R\left(f^{(2)}\right) + \frac{R(k)}{nh}$$

of the kernel density estimator

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

With this information, answer the following:

(a) Define (both in words and equations) $R(k)$, $\kappa_2^2(k)$ and $R(f^{(2)})$. Which are in the researchers control? Which are not in the researchers control?

(b) Derive the optimal bandwidth h_{opt}. Discuss what happens to this optimal bandwidth in large samples.

(c) If the AMISE of the first derivative of the density estimator is given as

$$AMISE\left[\widehat{f}'(x)\right] = \frac{h^4\kappa_2^2(k)}{4}R\left(f^{(3)}\right) + \frac{R(k')}{nh^3},$$

where k' is the first derivative of the kernel function with respect to x, what additional assumptions are we making? Derive the optimal bandwidth h_{opt} in this setting. Discuss what happens to this optimal bandwidth in large samples.

(a) $R(k) = \int k(y)^2 \, dy$

roughness of the kernel function

$K_2^2(k) = \left[\int y^2 k(y) \, dy \right]^2$

$K_2(k)$ is the second moment of the kernel function

$R(f^{(2)}) = \int f^{(2)}(x)^2 \, dx$

roughness of second derivative of the density

(b) $$\frac{\partial \, \text{AMSE}}{\partial h} = h^3 K_2^2(k) R(f^{(2)}) - \frac{f(x) R(k)}{n h^2} = 0$$

$$h^5 K_2^2(k) R(f^{(2)}) - \frac{f(x) R(k)}{n} = 0$$

$$h^5 = \frac{f(x) R(k)}{n K_2^2(k) R(f^{(2)})}$$

$$h = \left[\frac{f(x) R(k)}{K_2^2(k) R(f^{(2)})} \right]^{1/5} n^{-\frac{1}{5}}$$

as $n \to \infty$, $h \to 0$, $nh \to \infty$ which $h \to 0$ at rate $n^{-\frac{1}{5}}$

(c) We now assume an additional derivative on $f(x)$ & now have to pick a kernel w/ a first derivative

$$\frac{\partial \text{AIMSE}}{\partial h} = 0 \Rightarrow$$

$$h_{opt} = \left[\frac{3R(K')}{R(f^{(3)})\, k_2^2 C_k}\right]^{\frac{1}{7}} n^{-\frac{1}{7}}$$

as $n \to \infty$, $h \to 0$ & $nh \to \infty$

$h \to 0$ at rate $n^{-\frac{1}{7}}$

2. Suppose we were interested in a test for independence $f_{X,Y}(x,y) = f_X(x)f_Y(y)$. For simplicity, consider a single dimension for each random variable. With this information, answer the following:

   (a) Give a formula for the integrated square error. Show that this is equal to 0 under the null hypothesis.

   (b) Derive a feasible version of the integrated square error using leave-one-out estimators.

   (c) Discuss how you would bootstrap the distribution under the null hypothesis.

(a) $ISE = \int_y \int_x [f(x,y) - f(x)f(y)]^2 dxdy$

$H_0 : f(x,y) = f(x)f(y) \Rightarrow$

$ISE = \int_y \int_x [f(x)f(y) - f(x)f(y))]^2 dxdy$

$= \int_y \int_x (0)^2 dxdy = 0$

(b) $ISE = \int_y \int_x [f(x,y) \; f(x)f(y)]^2 dxdy$

$= \int_x \int_y f(x,y) dF(x,y) + \int_x f(x) dF(x)$

$\int_y f(y) dF(y)$

$-2 \int_y \int_x f(x) f(y) dF(x,y)$

$$\widehat{ISE}_n = \frac{1}{n}\sum_i \hat f_{-i}(x_i, y_i) + \frac{1}{n}\sum_i \hat f_{-i}(x_i)$$

$$\frac{1}{n}\sum_i \hat f_{-i}(y_i) - \frac{2}{n}\sum_i \hat f_{-i}(x_i)\hat f_{-i}(y_i)$$

$$= \frac{1}{n(n-1)h_x h_y}\sum_i \sum_j k\left(\frac{x_i - x_j}{h_x}\right) k\left(\frac{y_i - y_j}{h_y}\right)$$

$$+ \frac{1}{n^2(n-1)^2 h_x h_y}\sum_i \sum_{j\neq i} k\left(\frac{x_i - x_j}{h_x}\right)\sum_r \sum_{l\neq r} k\left(\frac{y_r - y_l}{h_y}\right)$$

$$- \frac{2}{n(n-1)^2 h_x h_y}\sum_i \sum_{j\neq i}\sum_{l\neq j} k\left(\frac{x_i - x_j}{h_x}\right) k\left(\frac{y_j - y_l}{h_y}\right)$$

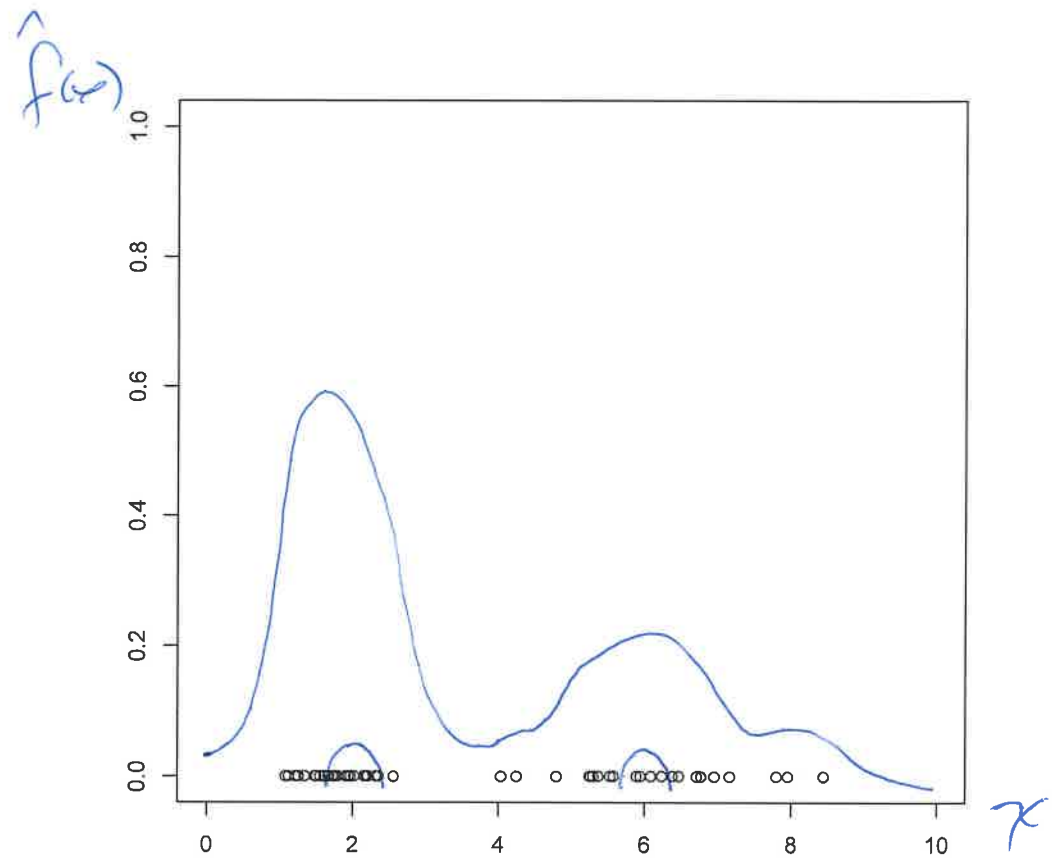
(c) resample from $\{x_i\}$ & $\{y_i\}$ separately, not in pairs, this does not change $\hat f(x)$ or $\hat f(y)$, but it does potentially change $\hat f(x,y)$, but ensures they are equal

3. Below is a set of 50 real observations drawn at random from an unknown density ($f(x)$). Using this figure, answer the following:

   (a) Write down a reasonable density that could have produced these observations. mixed kernel

   (b) For any two values of x, plot an Epanechnikov kernel function. What weight is given to each observation? $k_1(u) = \frac{3}{4}(1-u^2)\mathbb{1}\{|u| \le 1\}$

   (c) Label the axes and draw a reasonble density estimate ($\hat{f}(x)$).



(a) this was a mixture of normals

$N(2,1)$ w/ prob $\frac{1}{2}$

$N(6,2)$ w/ prob $\frac{1}{2}$

(b) for points "near" 2 the weight is $\frac{3}{4}(1-\varphi^2)$ where $\varphi = \left(\frac{x_i - 2}{h}\right)$ & weight 0 otherwise

Similar to "near" 6

(c) See figure on previous page