Econ 675: HW 2

Erin Markiewitz

October 11, 2018

Contents

1 Kernel Density Estimation

1.1

First we consider the kernel density derivative estimator. $\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{N} k\left(\frac{X_i - x}{h}\right)$

The expectation of the estimator is:

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

where $k^{(s)}$ is the s^{th} derivative of the kernel function Now integrate by parts

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

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

As $k(.)$ is a P^{th} order kernel function and $s - 1 < P$, the first term on the RHS of the equation above is equal to zero. Integrating by parts s-1 more times and changing the base, we get the following expression

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

So now we take a P^{th} order taylor expansion of $f^{(s)}(uh + x)$ around x, which gives us

$$f^{(s)}(x) + \frac{1}{P!}\int_{-\infty}^{\infty} k\left(u\right) f^{(s+P)}(uh + x)(uh + x - x)^p du + o(h_n^P)$$

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

$$= f^{(s)}(x) + \frac{f^{(s+P)}(x)}{P!}\mu_P(K)h_n^p + o(h_n^p)$$

where $\mu_P(K) = \int_{\mathbb{R}} u^P K(u)du$ - which gives the result. (Note: the second term is the bias of the estimator)

Now consider the variance of the estimator

$$\mathbb{V}[\hat{f}^{(s)}(x)] = \frac{1}{nh^{2+2s}}\mathbb{V}\left[[k^{(s)}\left(\frac{z-x}{h}\right)\right] = \frac{1}{nh^{2+2s}}\mathbb{E}\left[k^{(s)}\left(\frac{z-x}{h}\right)\right]^2 - \frac{1}{n}\mathbb{E}\left[\frac{1}{nh^{1+s}}k^{(s)}\left(\frac{z-x}{h}\right)\right]^2$$

Now using our derivation of the expected value of our estimator we can rewrite the expression above as:

$$\frac{1}{nh^{2+2s}}\mathbb{E}\left[k^{(s)}\left(\frac{z-x}{h}\right)\right]^2 - \frac{1}{n}f^{(s)}(x)^2 + O\left(\frac{1}{n}\right)$$

(This comes from $\{\frac{f^{(s+P)}(x)}{P!}\mu_P(K)h_n^p + o(h_n^p)\}$ being bounded)
So continuing on, we just expand the first term a bit

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

$$= \frac{1}{nh^{1+2s}}\int_{-\infty}^{\infty} k^{(s)}\left(u\right) f(uh+x)du - \frac{1}{n}f^{(s)}(x)^2 + O\left(\frac{1}{n}\right)$$

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

$$= \frac{f(x)\nu_s(k)}{nh^{1+2s}} - \frac{1}{n}f^{(s)}(x)^2 + O\left(\frac{1}{n}\right)$$

where $\nu_s(k) = \int_{\mathbb{R}} k^{(s)}\left(u\right)^2 du$ is the roughness of the s^{th} derivative of a given function k - which gives the result.

1.2

The optimal bandwith estimator solves the following problem

$$\min_h AIMSE[h] = \min_h \int_{-\infty}^{\infty}\left[\left(h_n^p\mu_p(k)\frac{f^{(P+s)}(x)}{P!}\right)^2 + \frac{\nu_s(k)f(x)}{nh_n^{1+2s}}\right]dx$$

Take first order conditions

$$0 = 2Ph^{2P-1}\int_{-\infty}^{\infty}\left[\left(\mu_p(k)\frac{f^{(P+s)}(x)}{P!}\right)^2 - \frac{(1+2s)\nu_s(k)f(x)}{nh^{2s}}\right]dx$$

$$\frac{2Pnh^{1-2P-2s}}{(1+2s)\nu_s(k)} = \left(\frac{P!}{\mu_p(k)\nu_{(P+s)}(f))}\right)^2$$
$$h_{AIMSE,s} = \left(\frac{(1+2s)\nu_s(k)(P!)^2}{2Pn\mu_p(k)^2\nu_{(P+s)}(f))}\right)^{\frac{1}{1-2P-2s}}$$
$$h_{AIMSE,s} = \left(\frac{(1+2s)(P!)^2}{2Pn}\frac{\nu_s(k)}{\mu_p(k)^2\nu_{(P+s)}(f))}\right)^{\frac{1}{1-2P-2s}}$$

Now for a consistent bandwith estimator we use cross validation procedure from the lecture notes. Cross-Validation minimizes the estimated mean-squared error through a choice of bandwith.

$$h^* = \text{argmin}_{h\in\mathbb{R}^{++}}\ CV(h) = \frac{1}{n^2h}\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{X_i - X_j}{h}\right)K\left(\frac{X_i - X_j}{h}\right) - \frac{2}{n}\sum_{i=1}^{n}\hat{f}_{-i}(X_i)$$

where $\hat{f}_{-i}(X_i)$ is the estimated density w/o x_i in the sample.

1.3

1.3.1

PUT GRAPHS HERE

1.3.2

Talk about convergence here!!!!

1.3.3 C

onsidering a rule-of-thumbs estimate of the bandwith, we assume the DGP is gaussian, so

$$\bar{h}_{AIMSE} = M^{-1}\sum_{m=1}^{M}\hat{h}_{AIMSE,m} =$$

2 Linear Smoothing, Cross-Validation and Series

2.1

Local polynomial regression solves the following problem:

$$\hat{\beta}_{LPR} = \text{argmin}_{\beta \in \mathbb{R}^P+1} \frac{1}{n} \sum_{i=1}^{N} (Y_i - r_p(x - x)\beta)^2 K(\frac{x_i - x}{h})$$

where $r_p(u) = (1, u, u^2, ..., u^p)'$ The true regression function $e(x_i)$ is estimated by $\hat{e}(x) = \hat{\beta}_{LPR}$, which can be rewritten as a weighted least-squares problem where $\hat{\beta}_{LPR}(x) = (\mathbf{R_p'WR_p})^{\mathbf{-1}}\mathbf{R_p'WY}$ where the weighting matrix is a diagonal matrix with the kernel functions of the x_i.

where

$$\mathbf{R}_p = \begin{bmatrix} 1 & (x_1 - x) & (x_1 - x)^2 & \cdots & (x_1 - x)^p \\ \vdots & \cdots & \ddots & \ddots & \vdots \\ 1 & (x_n - x) & \cdots & \cdots & (x_n - x)^p \end{bmatrix}$$

and $\mathbf{W}$ is a matrix with kernel weights of x_is on the diagonal

$$\mathbf{W} = \begin{bmatrix} K(\frac{x_1 - x}{h}) & 0 & 0 & 0 \\ 0 & K(\frac{x_2 - x}{h}) & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & K(\frac{x_n - x}{h}) \end{bmatrix}$$

So we can rewrite the estimator of our regression equation as
$\hat{e}(x) = \mathbf{e_1'}\hat{beta}_{LPR} = R_p'WR_p)^{-1}R_p'WY$
where $\mathbf{e_1}$ is a basis vector of length $1 + p$.
Therefore we can rewrite the estimator above as a sum.

$$\hat{e}(x) = \mathbf{e_1'} \left(\sum_{i=1}^{n} r_p(x_i - x) r_p(x_i - x)' K \left(\frac{x_i - x}{h} \right) \right)^{-1} \left(\sum_{i=1}^{n} r_p(x_i - x) r_p(x_i - x) y_i K \left(\frac{x_i - x}{h} \right) \right)$$

Now we consider the series estimator, which solves the following problem

$$\hat{\beta}_s = \text{argmin}_{\beta \in \mathbb{R}^{k_n}} \frac{1}{n} \sum_{i=1}^{N} (Y_i - r_{k_n}(x)\beta)^2 K(\frac{x_i - x}{h})$$

where $r_{k_n}(x)$ is the basis of some series defined on x, so that

$$\hat{e}(x) = \mathbf{r_{k_n}(x)}'\hat{\beta}$$

where

$$\hat{\mathbf{beta}_\mathbf{s}} = \left(\mathbf{R_p'R_p}\right)^{\mathbf{-1}} \mathbf{R_p Y}$$

and

$$\mathbf{R}_p = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^p \\ \vdots & \cdots & \ddots & \ddots & \vdots \\ 1 & x_n & \cdots & \cdots & x_n^p \end{bmatrix}$$

So we can rewrite the estimated regression function as

$$\hat{e}(x) = \mathbf{r_p(x)}' \left(\mathbf{R_p'R_p}\right)^{\mathbf{-1}} \mathbf{R_p Y}$$

and

$$\hat{e}(x) = \mathbf{r_p(x)}' \left(\sum_{i=1}^{n} r_p(x_i) r_p(x_i)'\right)^{-1} \left(\sum_{i=1}^{n} r_p(x_i) y_i\right)$$

2.2

Next, we need to show the following simplified cross-validation formula holds for local polynomial regression and series estimation:

$$CV(c) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{e}(x_i))^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{e}_{(i)}(x_i)}{1 - w_{n,1}(x_i)}\right)^2$$

where c is a tuning parameter (h_n for LPR or a truncation K for series estimators)

Note from the first part of the question, we found that we can right both regressor estimators as a weighted average of the outcome variable

$$\hat{e}(x) = \frac{1}{n}\sum_{i=1}^{n} w_{n,i}(x) y_i$$

where $w_{n,i}(x) = w_{n,i}(x_1, x_s, \cdots, x_n; x)$

Now in estimation of the tuning parameter, we need our smoothing parameter $w_{n,i}$ to be consistent when we "leave one (x_i) out" for estimation. Our smoothing parameter sums to one in the case of the LPR and series estimators. So for cross validation, we need to adjust accordingly.

$$\hat{e}_{(i)}(x) = \frac{1}{1 - w_{ii}} \sum_{i=1}^{n} w_{i,j}(x) y_i$$

So to get our result:

$$\begin{aligned}
(1 - w_{ii})\hat{e}_{(i)}(x) &= \sum_{j \neq i, j=1}^{n} w_{i,j}(x) y_j \\
\hat{e}_{(i)}(x) &= \sum_{j \neq i, j=1}^{n} w_{i,j}(x) y_j + w_{i,i}\hat{e}_{(i)}(x) \\
&= \sum_{i=1}^{n} w_{i,j}(x) y_i + w_{ii}\hat{e}_{(i)}(x) - w_{i,i} y_i \\
&= \hat{e}(x) + w_{ii}\hat{e}_{(i)}(x) - w_{i,i} y_i
\end{aligned}$$

Which gives us

$$\begin{aligned}
y_i - \hat{e}_{(i)}(x) &= y_i - \hat{e}_(x) - w_{i,i}\hat{e}_{(i)}(x) + w_{i,i} y_i \\
y_i - \hat{e}_{(i)}(x) &= y_i - \hat{e}_(x) w_{i,i}(y_i - \hat{e}_{(i)}(x)) \\
(1 - w_{i,i})(y_i - \hat{e}_{(i)}(x)) &= y_i - \hat{e}_(x) \\
y_i - \hat{e}_{(i)}(x) &= \frac{y_i - \hat{e}_(x)}{1 - w_{i,i}}
\end{aligned}$$

So it follows

$$\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{e}(x_i))^2 = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_i - \hat{e}_{(i)}(x_i)}{1 - w_{n,1}(x_i)} \right)^2$$

2.3

If we assume the data is iid and a finite first moment, we can show consistency:

$$\begin{aligned}
\mathbb{E}[\hat{e}(x)|x] &= \mathbb{E}\left[\sum_{i=1}^{n} w_{n,i}(x_i) y_i \right] \\
&= \sum_{i=1}^{n} w_{n,i}(x_i) \mathbb{E}[y_i|x] \\
&= \mathbb{E}[y_i|x]
\end{aligned}$$

as $\sum_{i=1}^{n} w_{n,i}(x_i) = 1$

Now if we assume a finite second moment of our regressor estimator

$$\begin{aligned}
\mathbb{V}[\hat{e}(x)|x] &= \mathbb{V}\left[\sum_{i=1}^{n} w_{n,i}(x_i)y_i\right] \\
&= \sum_{i=1}^{n} \mathbb{V}[w_{n,i}(x_i)y_i|x] \\
&= \sum_{i=1}^{n} w_{n,i}(x_i)^2 \mathbb{V}[y_i|x] \\
&= \mathbb{V}[y_i|x] \sum_{i=1}^{n} w_{n,i}(x_i)^2
\end{aligned}$$

So the variance estimator is

$$\hat{V}(x) = \left(\frac{1}{1-n}\sum_{i=1}^{n}(y_i - \hat{e}(x_i))^2\right)\left(\sum_{i=1}^{n} w_{n,i}(x_i)^2\right)$$

Asymptotic normality follows from the CLT

2.4

The

A pointwise asymptotically valid confidence interval for a fixed x is

$$CI_{95} = \left[\hat{e}(x) - 1.96\sqrt{\hat{V}(x)/n}; \hat{e}(x) + 1.96\sqrt{\hat{V}(x)/n}\right]$$

but this convidence invterval is not asymptotically valid across the entire support. In order for the confidence interval to be uniformly aymptotically valid we need the interval to hold across the entire support of x

$$\sup_{x\in X}\left|\frac{\hat{e}(x) - e(x))}{\sqrt{\hat{V}(x)}}\right| \leq q_{1-\alpha/2}$$

3 Semiparametric Semi-Linear Model

3.1

The following question concerns this moment condition:

$$\mathbb{E}[(t_i - h_0(x_i))(y_i - t_i\theta)] = 0 \text{ , where } h_0(x_i) = \mathbb{E}[t_i|x_i]$$

As long as t_i is not collinear with x_i then θ_0 will be identifiable. Assuming that θ_0 is identifiable, it satisfies the moment condition above:

$$\begin{gathered}
\mathbb{E}[t_i y_i] + \mathbb{E}[h_0(x_i)t_i\theta] - \mathbb{E}[h_0(x_i)y_i] - \mathbb{E}[t_i t_i\theta)] = 0 \\
\mathbb{E}[\mathbb{E}[t_i y_i|t_i, x_i]] + \mathbb{E}[\mathbb{E}[h_0(x_i)t_i\theta|t_i, x_i]] - \mathbb{E}[\mathbb{E}[h_0(x_i)y_i|t_i, x_i]] + \mathbb{E}[\mathbb{E}[t_i t_i\theta)|t_i, x_i]] = 0 \\
\mathbb{E}[h_0(x_i)\mathbb{E}[y_i|t_i, x_i]] + \mathbb{E}[h_0(x_i)h_0(x_i)]\theta - \mathbb{E}[h_0(x_i)\mathbb{E}[y_i|t_i, x_i]] + \mathbb{E}[h_0(x_i)h_0(x_i)]\theta = 0 \\
0 = 0
\end{gathered}$$

To derive a closed form equation for θ_0 we follow the steps outlined in Hansen's notes on nonparametrics (chapter 7), which describes Robinson (Econometrica, 1988).

$$y_i = t_i\theta_0 + g(x_i) + \epsilon_i$$

First we take the conditional expectation with respect to the treatment and other covariates. (We assume the treatment is not collinear with the other covariates.)

$$\mathbb{E}[y_i|t_i x_i] = \mathbb{E}[t_i|t_i x_i]\theta_0 + \mathbb{E}[g(x_i)|t_i x_i] + 0\mathbb{E}[y_i|t_i x_i] = h_o(x_i)\theta_0 + g(x_i) + 0$$

Next, let's define $g_{y,x} \coloneqq \mathbb{E}[y_i|t_i x_i]$, and subtract the equation above from the original regression.

$$y_i - g_{y,x} = (t_i - h_o(x_i))\theta_0 + g(x_i) - g(x_i) + \epsilon_i$$

Now, we can rewrite the regression as a residual regression:

$$\begin{gathered}
\epsilon_{yi} = \epsilon_{ti}\theta_0 + \epsilon_i \\
y_i = g_{y,x} + \epsilon_{yi} \\
t_i = h_o(x_i) + \epsilon_{ti}
\end{gathered}$$

Which produces the infeasible estimtor:

$$\beta = \left(\sum_{i=1}^{n} \epsilon_{ti}\epsilon_{ti}'\right)^{-1} \left(\sum_{i=1}^{n} \epsilon_{ti}\epsilon_{yi}'\right)$$

Note that we can rewrite the residual regression as :

$$M_{yx}y_i = M_{tx}t_i\theta_0 + \epsilon_i$$

Which is the second stage of an IV regression that partials out the effects of X_i on y_i and t_i using anhilation matrixes.

3.2

3.2.1

If the treatment is undetermined by the power series of the covariates, θ_0 is simply

$$\theta_0 = (T'T)^{-1}(T'Y)$$

which has a feasible estimator of

$$\hat{\theta}(K) = (\sum_{i=1}^{n} t_i t_i)^{-1}(\sum_{i=1}^{n} t_i y_i)$$

3.2.2

If the treatment is correlated to the other covariates, in order to estimate a feasible estimator, one must run Nadaraya - Watson kernel regressions of the outcome and treatment variables onto the power series.

$$\hat{y}_i = \frac{\sum_{i=1}^{n} k\left(\frac{p^{K_n}(x_i) - p^{K_n}(x)}{h}\right) y_i}{\sum_{i=1}^{n} k\left(\frac{p^{K_n}(x_i) - p^{K_n}(x)}{h}\right)}$$

$$h_0(x_i) = \frac{\sum_{i=1}^{n} k\left(\frac{p^{K_n}(x_i) - p^{K_n}(x)}{h}\right) t_i}{\sum_{i=1}^{n} k\left(\frac{p^{K_n}(x_i) - p^{K_n}(x)}{h}\right)}$$

Now, construct residualize

$$\hat{\epsilon}_{yi} = y_i - \hat{y}_i = M_{yx}y_i$$
$$\hat{\epsilon}_{ti} = t_i - h_0(x_i) = M_{tx}t_i$$

Which produces the feasible estimator

$$\hat{\theta}(K) = \left(\sum_{i=1}^{n} \hat{\epsilon}_{ti}\hat{\epsilon}'_{ti}\right)^{-1} \left(\sum_{i=1}^{n} \hat{\epsilon}_{ti}\hat{\epsilon}'_{yi}\right)$$

3.3

3.3.1

Fixing K, the reason this approach is called a "flexible parametric" estimation because you are estimating θ_0, while letting

If $K \to \infty$ does not invalidate the "fixed K" assumption as long as the ratio between the observations and covariates is fixed $\left(\frac{K_n}{n} = \frac{\bar{K}}{\bar{n}}\right)$

3.3.2

Using the results above the confidence interval is

$$CI_{95} = \left[\hat{\theta}(K) - 1.96\sqrt{\hat{V}_{HCO}/n}; \hat{\theta}(K) + 1.96\sqrt{\hat{V}_{HCO}/n}\right]$$

3.4