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

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

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

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

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

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

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

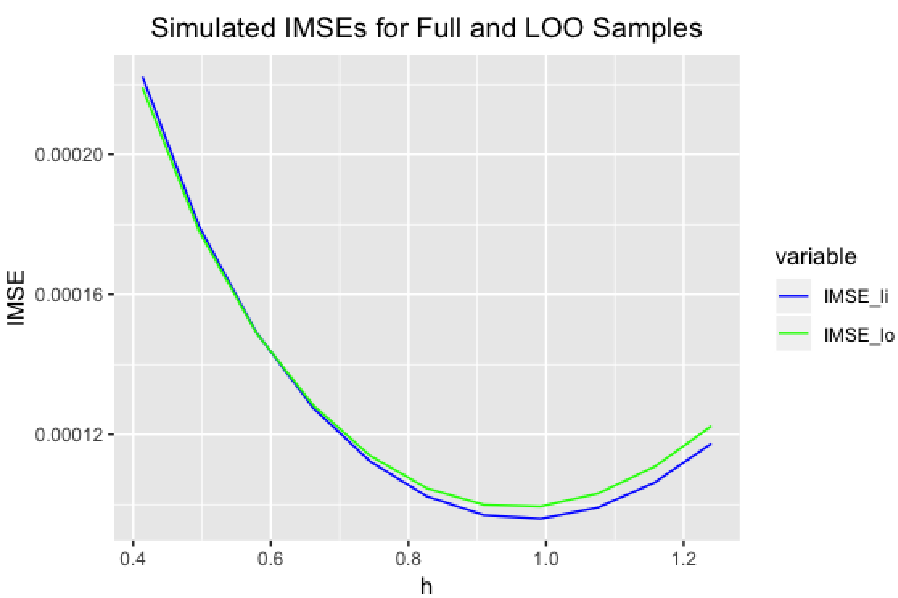
Now for a consistent bandwith estimator we use cross validation procedure from the lecture notes. Cross-Validation minimizes the estimated mean-squared error through a choice of bandwith.

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

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

1.3

1.3.1



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:

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

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 $\hat{\beta}_{LPR}(x) = (\mathbf{R_p' W R_p})^{\mathbf{-1}} \mathbf{R_p' W Y}$ 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
$\hat{e}(x) = \mathbf{e_1'} \hat{beta}_{LPR} = R_p' W R_p)^{-1} R_p' W Y$
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

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

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

$$\hat{\mathbf{beta_s}} = \left(\mathbf{R_p'R_p}\right)^{-1}\mathbf{R_pY}$$

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

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

and

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

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:

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