Lecture Note 14.771: Kernel Regression

Esther Duflo

September 2002

$$y = g(x) + \epsilon$$

Problem is to estimate $g(x)$ without imposing a functional form.

- Series estimator: regress y on polynomials of x, splines, etc...
- Common non-parametric estimator = the kernel estimator

$g(x)$ is the conditional expectation of y given x.

$$g(x) = E(y|x) = \int yf(y|x)dy$$

By Bayes's rule:

$$\int yf(y|x)dy = \int \frac{yf(x,y)dy}{f(x)} = \frac{\int yf(x,y)dy}{f(x)}$$

The kernel estimator replaces $yf(x,y)$ and $f(x)$ by their empirical estimates.

$$\widehat{g}(x) = \frac{\int y\widehat{f}(x,y)dy}{\widehat{f}(x)}$$

- Denominator: estimating the density of x.

$$\frac{1}{N*h}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right),$$

where h is a positive number (the bandwidth), is the kernel estimate of the density of x. $K(.)$ is a density, i.e. a positive function that integrates to 1.

It is a weighted proportion of observations that are within a distance h of the point x.

Examples

1. Histogram: $K(u) = 1/2$ if $|u| \leq 1, K(u) = 0$ otherwise.

2. Epanechnikov: $K(u) = \frac{3}{4}(1-u^2)$ if $|u| \leq 1, K(u) = 0$ otherwise.

3. Quartic: $K(u) = [\frac{3}{4}(1-u^2)]^2$ if $|u| \leq 1, K(u) = 0$ otherwise.

- Numerator:

$$\frac{1}{N*h}\sum_{i=1}^{n} y_i K\left(\frac{x-x_i}{h}\right)$$

Proof (Bivariate kernel estimator):

$$\widehat{f}(x,y) = \frac{1}{N}\sum_{i=1}^{n}\frac{1}{h^2}\widetilde{K}\left(\frac{x-x_i}{h}, \frac{y-y_i}{h}\right)$$

where

$$\left| \begin{array}{c} K(u,v) = K(-u,-v) \\ \int\int \widetilde{K}(u,v)dudv = 1 \\ \int \widetilde{K}(u,v)du = K(v) \\ \int v\widetilde{K}(u,v)dv = 0 \end{array} \right.$$

$$\begin{aligned}
\int y\widehat{f}(x,y)dy &= \int y\frac{1}{N}\sum_{i=1}^{n}\frac{1}{h^2}\widetilde{K}\left(\frac{x-x_i}{h}, \frac{y-y_i}{h}\right)dy \\
&= \frac{1}{N}\sum_{i=1}^{n}\int\frac{1}{h^2}y\widetilde{K}\left(\frac{x-x_i}{h}, \frac{y-y_i}{h}\right)dy \\
&= \frac{1}{N}\sum_{i=1}^{n}\frac{1}{h}\int(y_i+hv)\widetilde{K}\left(\frac{x-x_i}{h}, v\right)dv \\
&= \frac{1}{N}\sum_{i=1}^{n}y_i\int\frac{1}{h}\widetilde{K}\left(\frac{x-x_i}{h}, v\right)dv + h\int\frac{1}{h}v\widetilde{K}\left(\frac{x-x_i}{h}, v\right)dv \\
&= \frac{1}{N*h}\sum_{i=1}^{n}y_i K\left(\frac{x-x_i}{h}\right)
\end{aligned}$$

In summary:

$$\widehat{g}(x) = \frac{\sum_{i=1}^{n} y_i K\left(\frac{x-x_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)} \tag{1}$$

$\widehat{g}(x)$ is a weighted average of y over a range close to x. The weights are declining for points further away from x.

In practice, you choose a grid of points (ex. 50 points) and you calculate the formula given in equation 1 for each of these points.

Exercise: Alternative: Locally weighted regression (Fan, 92). At each point, calculate a weighted regression of y on x and a constant, using the kernel weights: $K\left(\frac{x-x_i}{h}\right) = w_i$. Run $\sqrt{w_i}y = \alpha\sqrt{w_i}x + b + \in_i$, and form

$$\widehat{g}(x) = \widehat{y}.$$

Obtain formula very close to the kernel formulas, but better at the boundaries, or anywhere where there is a disconituity. (See Deaton's book).