Lecture Note 12: Non-parametric Methods

NOT ON FINAL.

parametric vs. non-parametric
- ① estimating the density
- ② estimating the reg. function

Linear relationship: $Y = \beta_0 + \beta_1 X + U$

$\downarrow$

$E[Y|X] = \beta_0 + \beta_1 X$

Non-linear relationship? $E[Y|X] = g(X)$

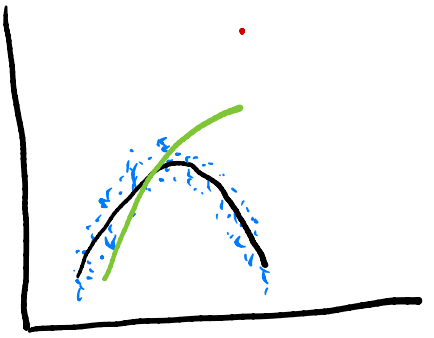
Want to estimate: $Y = g(X) + U$

↳ 2 methods: 1) polynomial regression
2) local linear regression

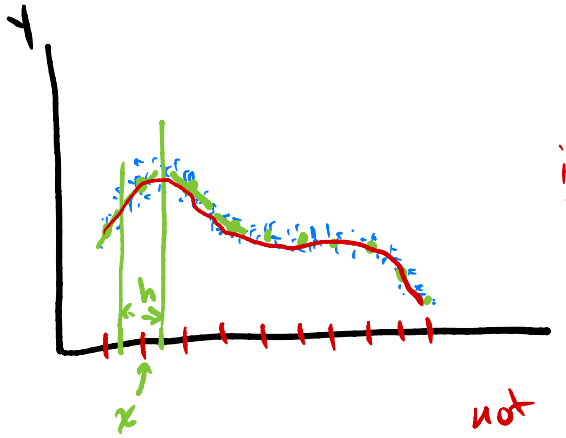
Polynomial regression

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \cdots + \beta_k X^k + U$$

- Calculus: As $k \to \infty$, polynomial converges to $g(X)$
- bias vs. variance: high k → less bias, more variance
  low k → more bias, less variance
- unappealing: outlier sensitivity



Local regression



→ local linear regression

→ bandwidth: h (half of window)

important

→ bias vs. variance → high h: more bias, less variance

low h: less bias, more var

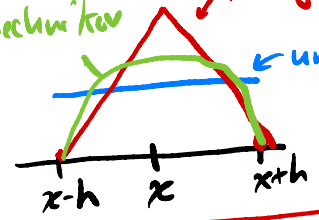
→ kernel function: $k\left(\frac{x_i - x}{h}\right)$

not that important

Epanechnikov

triangular

uniform (rectangular)



→ for each x, estimate:

$$\min_{b_0, b_1} \sum_i k\left(\frac{x_i - x}{h}\right)(Y_i - b_0 - b_1 X_i)^2$$

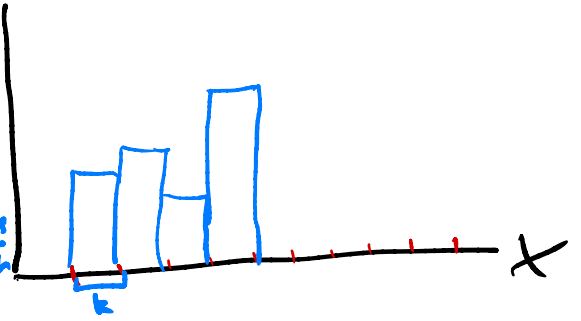
Stata: lpoly

R: KernSmooth::locpoly()

ggplot::geom_smooth()

Estimating the density

→ Histogram:
- → vertical axis: → count / frequency: N_k
  - → share of obs: N_k/N
  - → density: $\hat{f}(x) = \frac{N_k}{N} \cdot \frac{\text{num of bins}}{\text{bin width}}$



→ Stata: hist

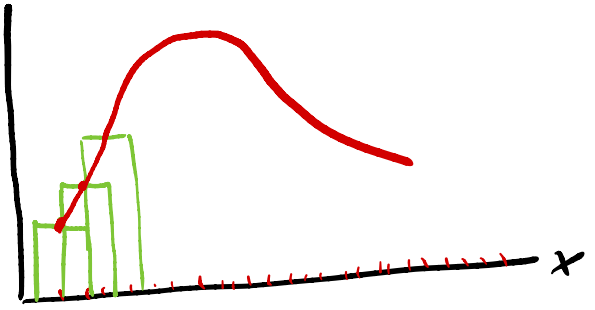
R: hist() or ggplot::geom_histogram()

→ Centered histogram

↳

→ kernel density estimation

$$\hat{f}(x) = \frac{1}{Nh} \sum_{i=1}^{N} k\left(\frac{x_i - x}{h}\right)$$



Stata: kdensity

R: ggplot::geom_density()