Agglomeration Economies

Urban Economics

Ignacio Sarmiento-Barbieri

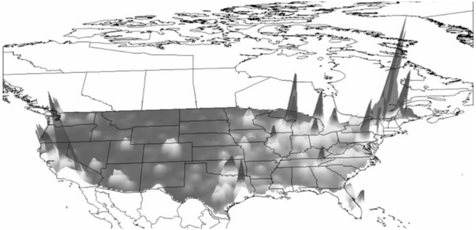
Universidad de los Andes

September 27, 2023

Agenda

Agglomeration Economies

- Why do we see such a remarkable clustering of human activity in a small number of urban areas?



Evidence of Agglomeration Economies

- Three strategies to identify agglomeration economies

  1. Show there is too much spatial concentration to be random (Duranton and Overman, 2005)
  2. Compare productivity over space (Greenstone, 2010)
  3. Compare wages and rents across space (Quantitative Spatial Models, Ahlfeldt et al, 2015)

Agenda

Spatial Concentration

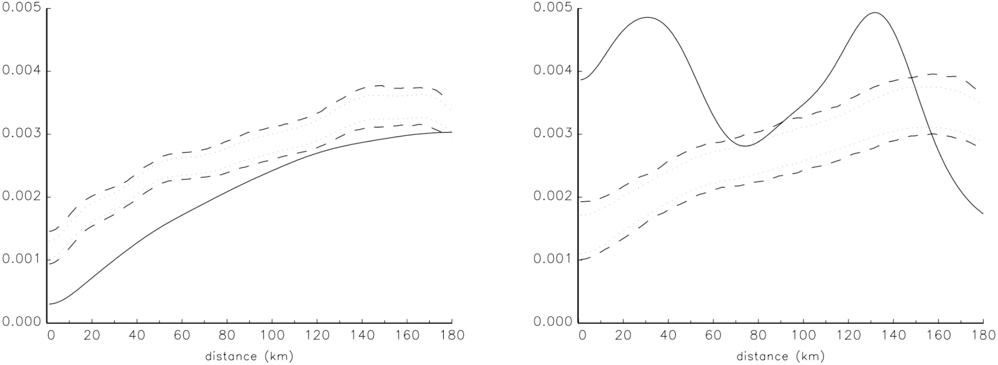
Extremes of Localization and Dispersion

(c) Other Agricultural and Forestry Machinery (SIC2932)

(d) Machinery for Textile, Apparel and Leather Production (SIC2954)

Spatial Concentration

K Density Estimates



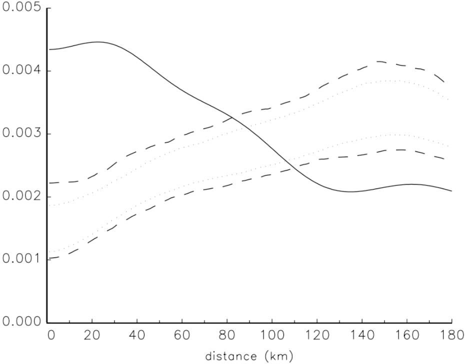
(c) Other Agricultural and Forestry Machinery (SIC2932)

(d) Machinery for Textile, Apparel and Leather Production (SIC2954)

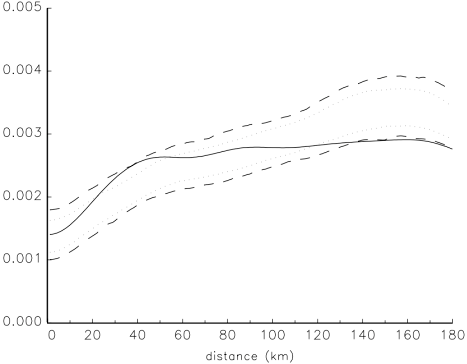
Figure 2. *K*-density, local confidence intervals and global confidence bands for four illustrative industries

Spatial Concentration

Ambiguous Cases



(a) Basic Pharmaceuticals (SIC2441)



(b) Pharmaceutical Preparations (SIC2442)

Agenda

Agenda

Univariate density estimation: parametric vs nonparametric methods

- Let $f = f(.)$ be the density function of the random variable X
- Let $x_1,\ x_2,\ \ldots x_n$ be a random sample of X. Then, $x_i \sim f\ iid$
- How can we estimate the density in a particular point x_0, then $f(x_0)$?

Agenda

Parametric methods

Parametric methods

- They assume a particular functional form for f. :

Parametric methods

- They assume a particular functional form for f. :
- Example:

$$f(x_o) = \phi(x_o) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x_o - \mu}{\sigma}\right)^2\right] \tag{1}$$

- The ignorance about $f(x_o)$ is limited to ignorance of the two parameters μ and σ

Parametric methods

- They assume a particular functional form for f. :
- Example:

$$f\left(x_o\right)=\phi\left(x_o\right)=\frac{1}{\sigma\sqrt{2\,\pi}}\exp\left[-\frac{1}{2}\left(\frac{x_o-\mu}{\sigma}\right)^2\right] \tag{1}$$

- The ignorance about $f\left(x_o\right)$ is limited to ignorance of the two parameters μ and σ
- Consistent estimators (ML) are:

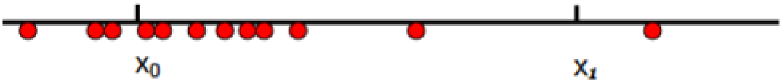
$$\hat{\mu} = \bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{2}$$

$$\hat{\sigma} = S^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}{n} \tag{3}$$

Agenda

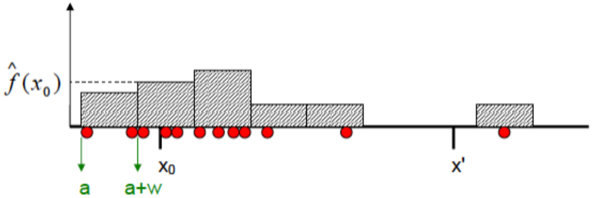
Non-parametric methods

- They seek to estimate $f(x_o))$ without assuming a particular functional form, only assuming certain regularity conditions of the density (smoothness, differentiability)
- How is the sample information interpreted?
- If we observe more data "near" x_o than x_1 we infer that $f(x_o) > f(x_1)$



A rudimentary non-parametric estimator: the histogram

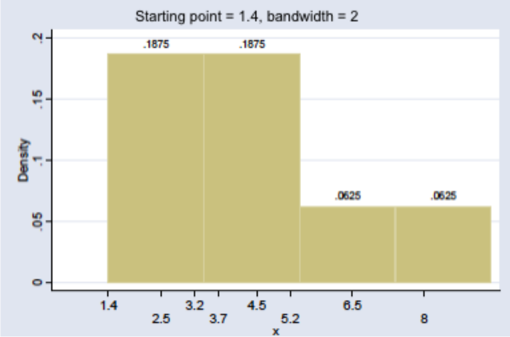
- It consists of estimating the probability within intervals through the relative frequency of observations within that interval.
- The intervals are determined from an initial point a and a bandwidth w



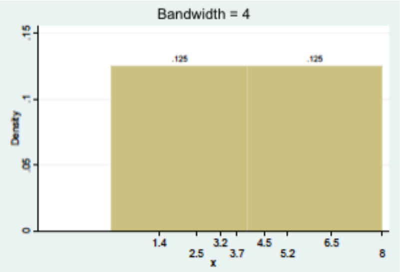
- The area of the bars is the relative frequency: $\frac{Nber\,obs.\,in\,interval}{n}$
- The height of the bars is an estimator of the density at any point in the interval

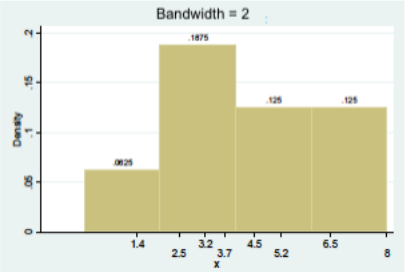
$$\hat{f}(x_o) = \frac{Nber.\,obs\ interval}{n \times w} \tag{4}$$

Histogram problems: (1) depends on starting point

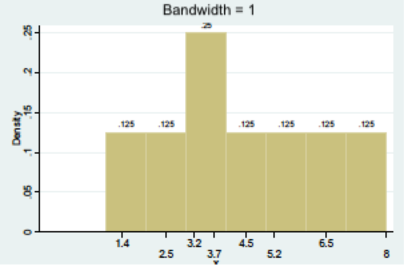


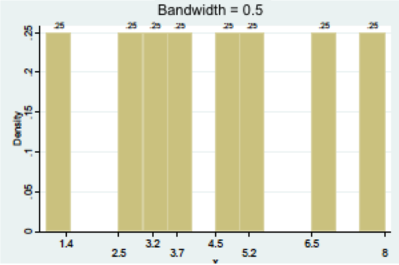
Histogram problems: (2) depends on the bandwidth





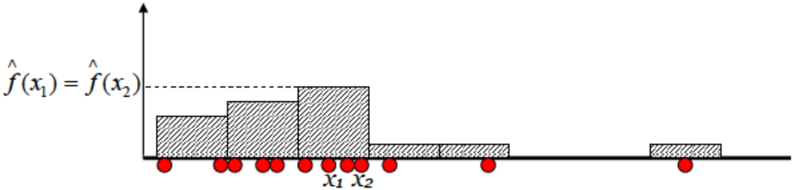
Histogram problems: (2) depends on the bandwidth





Histogram problems: (3) is discontinuous at the ends of the interval

- Note that $\hat{f}\ (x_1) = \hat{f}\ (x_2)$
- But $\hat{f}\ (x_2 + \epsilon) = \frac{1}{4}\hat{f}\ (x_2)$ for any $\epsilon > 0$



Another non-parametric estimator: the naive estimator

Discrete case

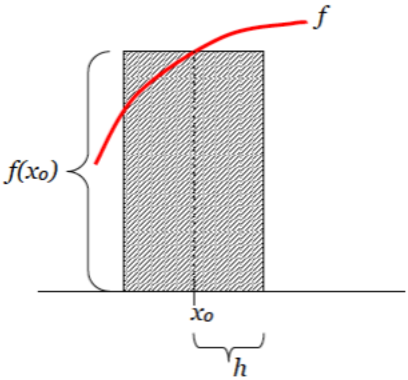
- X is a discrete VA, *iid*
- Objective: estimate $\Pr\left(X = x_o\right) = f\left(x_o\right)$
- The naive estimator comes from:

$$\hat{f}\left(x_o\right) = \frac{\# \, x_i = x_o}{n} = \frac{1}{n} \sum_{i=1}^{n} I\left(x_i = x_o\right) \tag{5}$$

Another non-parametric estimator: the naive estimator

Continuous case

- X is a continuous VA ($Pr\,(X = x0) = 0$) we evaluate the probability that X is "close" to x_0.
- We say that x_i is close to x_0 if x_i belongs to the interval $(x_o -\ h,\ x_o +\ h)$



Another non-parametric estimator: the naive estimator

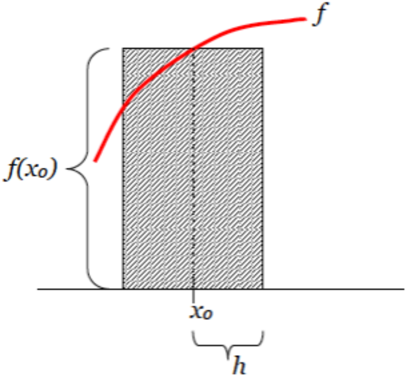
- Formally:

$$
\begin{aligned}
2h\ \times f\left(x_0\right) &\approx \int_{x_o-h}^{x_o+h} f\left(z\right) dz = \Pr\ \left[x_o - h\ < X < x_o + h\right] && (6)\\
&\approx \frac{\#\, x_i\ \epsilon\ \left(x_0 - h, x_o + h\right)}{n} && (7)\\
&= \frac{1}{n}\sum_{i=1}^{n} I\left(x_i\ \epsilon\ \left(x_0 - h, x_o + h\right)\right) && (8)\\
&= \frac{1}{n}\sum_{i=1}^{n} I\left(x_0 - h < x_i < x_o + h\right) && (9)\\
&= \frac{1}{n}\sum_{i=1}^{n} I\left(-1 < \frac{x_i - x_o}{h} < 1\right) && (10)\\
& && (11)
\end{aligned}
$$

Another non-parametric estimator: the naive estimator

Continuous case

- X is a continuous RV ($Pr\left(X = x_0\right) = 0$) we evaluate the probability that X is "close" to x_0.
- We say that x_i is close to x_0 if x_i belongs to the interval $(x_o - \ h,\ x_o + \ h)$



The naive estimator

- Then:

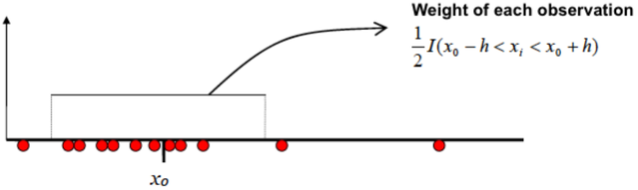
$$f(x_o) \cong \frac{1}{2h \times n} \sum_{i=1}^{n} I \left(-1 < \frac{x_i - x_o}{h} < 1 \right) \tag{12}$$

- $\hat{f}(x_o)$ is an estimator of the height of the rectangle in the previous graph

The naive estimator

Notice:

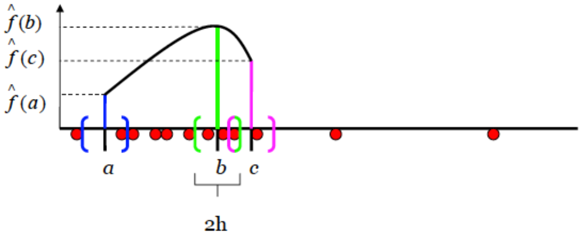
- Every possible point x_o is the center of an interval
- Observations that are within that interval (less than one h away from x_o) are weighted by 1⁄2



The naive estimator

Notice:

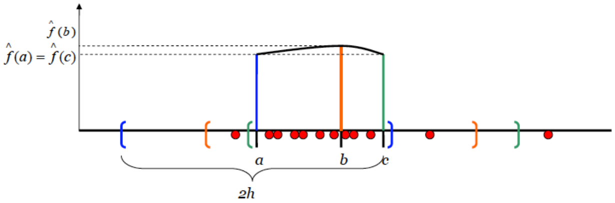
- To estimate the density at points a, b and c we construct intervals around them
- Then, unlike what happens in the histograms, the intervals of the naive estimator overlap (the initial point no longer matters)



The naive estimator

Problems

- It depends on h: the larger the bandwidth, the more distant observations from x_o are used to estimate f(x_o).The higher h, the smoother the estimated density (h=smoothing parameter)



- The weight $1/2I(.)$ is discontinuous at the limits of each interval, generating discontinuities in the estimated density.
- The weight treats observations very close to x_o in the same way as others somewhat further away, as long as they belong to the interval of length 2h around x_o

The weighted average estimator or kernel method

- The kernel estimator is a generalization of the naive estimator that overcomes some of the deficiencies of the latter.
- The weight $1/2I(.)$ of the naive estimator is replaced by a new weight K(.):

$$\hat{f}\ (x_o) = \frac{1}{n\ \times h} \sum_{i=1}^{n} K\left(\frac{x_i - x_0}{h}\right) \tag{13}$$

- Rosenblatt-Parzen density estimation approach

The weighted average estimator or kernel method

- The weights or kernels are known functions that satisfy:

1. $K(\phi) \geq 0$
2. $\int_{-\infty}^{\infty} K(\phi)\, d\phi = 1$
3. $K(\phi) = K(-\phi)$

The weighted average estimator or kernel method

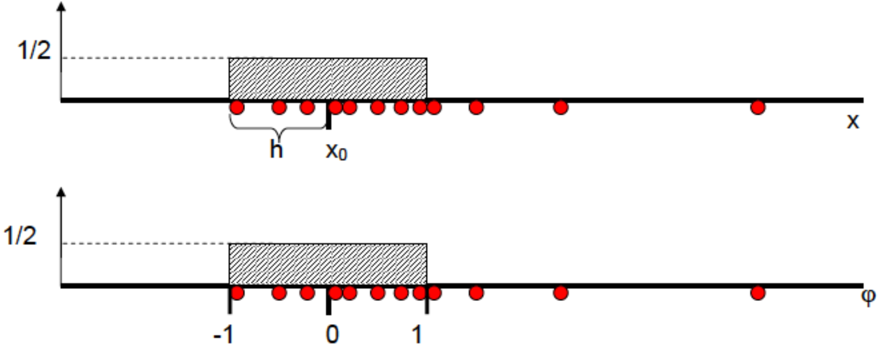
Example 1: rectangular or uniform kernel

$$K(\phi) = \begin{cases} \frac{1}{2} & |\phi| < 1 \\ 0 & o.w. \end{cases} \tag{14}$$

where $\phi = \frac{x_i - x_o}{h}$ In this case, the Kernels estimator matches the naive estimator. Note that the weight is similar to a uniform density function on the interval (-1, 1)

The weighted average estimator or kernel method

Example 1: rectangular or uniform kernel



The weighted average estimator or kernel method

Example 2: Gaussian kernel

- Gaussian

$$K(\phi) = \frac{1}{\sqrt{2\pi}} e^{-\phi^2/2} \tag{15}$$

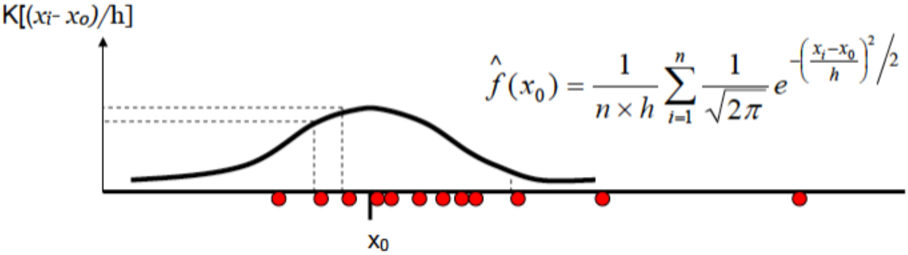
- The kernel function corresponds to the standard normal density function that satisfies the above assumptions.
- For this type of kernel, the density estimator is given by:

$$\hat{f}\,(x_o) = \frac{1}{n\ \times h} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\ \left(\frac{x_i - x_o}{h}\right)^2/2} \tag{16}$$

- Important: we don't assume normal distribution for any variable. The functional form is just used to weight the sample observations

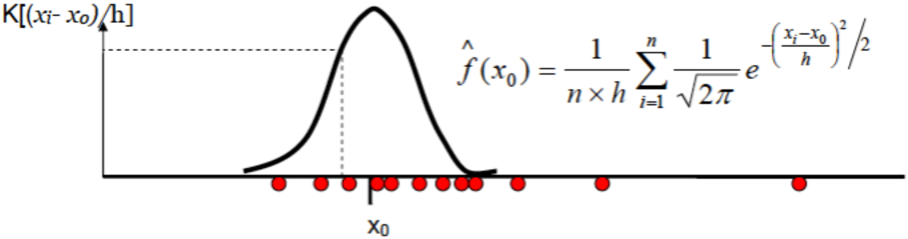
The weighted average estimator or kernel method

Example 2: Gaussian kernel



The weighted average estimator or kernel method

Example 2: Gaussian kernel (smaller h)



$$\hat{f}(x_0) = \frac{1}{n \times h} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{x_i - x_0}{h}\right)^2 / 2}$$

The weighted average estimator or kernel method

Other Kernels

Kernel type	Formula	Support
Gaussian or normal	$\frac{1}{\sqrt{2\pi}}\exp(-\varphi^2/2)$	$\varphi \in (-\infty, \infty)$
Epanechnikov	$\frac{3}{4\sqrt{5}}(1-\frac{1}{5}\varphi^2)I(\lvert\varphi\rvert<\sqrt{5})$	$\varphi \in \left(-\sqrt{5}, \sqrt{5}\right)$
Epanechnikov modificado	$\frac{3}{4}(1-\varphi^2)I(\lvert\varphi\rvert<1)$	$\varphi \in (-1, 1)$
Triangular	$(1-\lvert\varphi\rvert)I(\lvert\varphi\rvert<1)$	$\varphi \in (-1, 1)$
Uniform or rectangular	$\frac{1}{2}I(\lvert\varphi\rvert<1)$	$\varphi \in (-1, 1)$

The weighted average estimator or kernel method

Properties: Bias

- The kernel estimator is generally biased.
- The approximate expression for the bias is given by:

$$bias[\hat{f}(x_o)] \approx \frac{h^2}{2} f''(x_o) \int_{-\infty}^{\infty} K(\phi)\phi^2 d\phi \tag{17}$$

- The approximate expression for the asymptotic variance of the kernel estimator is given by:

$$variance[\hat{f}(x_o)] \approx \frac{1}{n \times h} f(x_o) \int_{-\infty}^{\infty} K^2(\phi) d\phi \tag{18}$$

The weighted average estimator or kernel method

Bandwidth Selection

$$MSE(\hat{f}(x)) = Var(\hat{f}(x)) + \left(bias(\hat{f}(x))\right)^2 \tag{19}$$

The weighted average estimator or kernel method

Bandwidth Selection

$$MSE(\hat{f}(x)) = Var(\hat{f}(x)) + \left(bias(\hat{f}(x))\right)^2 \tag{19}$$

- We obtain a bandwidth which globally balances bias and variance by minimizing MSE with respect to h, i.e.,

$$h_{opt} = \left(\frac{\int K^2(z)dz}{\left(\int z^2 K(z)dz\right)^2 \int f'(x)^2 dx}\right)^{-1/5} n^{-1/5} \tag{20}$$

The weighted average estimator or kernel method

Bandwidth Selection

- There are two popular approaches to bandwidth selection,
  1. Rules-of-thumb,

The weighted average estimator or kernel method

Rule-of-Thumb

- The rule-of-thumb for choosing the bandwidth makes assumtions about f and K
- For example: under a gaussian density and kernel

$$h_{opt} = 1.059\sigma n^{-1/5} \tag{21}$$

The weighted average estimator or kernel method

Bandwidth Selection

- There are two popular approaches to bandwidth selection,
  1. Rules-of-thumb,
  2. cross-validation methods,

Agenda

Spatial Concentration

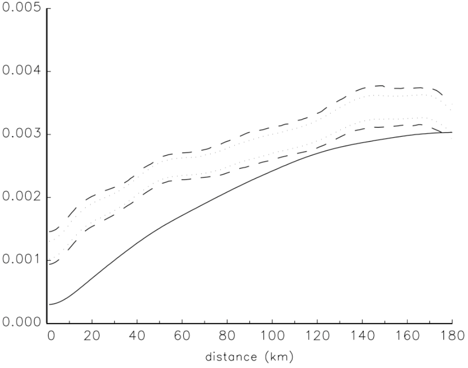
Extremes of Localization and Dispersion

(c) Other Agricultural and Forestry Machinery (SIC2932)

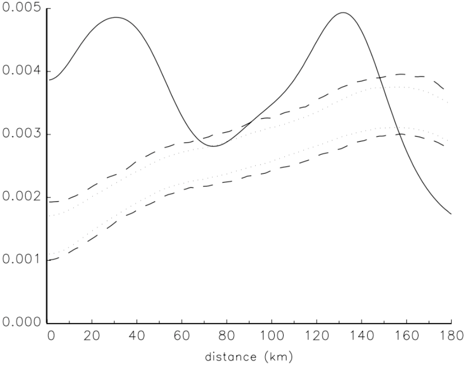
(d) Machinery for Textile, Apparel and Leather Production (SIC2954)

Spatial Concentration

K Density Estimates



(c) Other Agricultural and Forestry Machinery (SIC2932)



(d) Machinery for Textile, Apparel and Leather Production (SIC2954)

Figure 2. *K*-density, local confidence intervals and global confidence bands for four illustrative industries