Topics in Applied Econometrics for Public Policy

TA Session 1

Sergi Quintana

Barcelona School of Economics

April 19, 2023

TA Session 1

You can contact me at: sergi.quintana@bse.eu

Today's session

There will be two main modules:

- Revision and Stata implemetation of class material:
  - Histogram
  - Kernel density.
  - Choice of Kernel.
  - Chioce of Bandwidth
- Confidence interval estimation:
  - The bias problem.
  - Asymptotic confidence intervals.
  - Bootrap and jacknife confidence intervals.

Motivation

- Random variables are characterized by their cumulative distribution function (cdf).
- The problem is that cdfs are hard to visualize and interpret.
- For this reason, we will approximate probability density functions.
- Density functions provide intuitive information about the highest density regions, mode and shape of the support of a random variable.

Histogram

- The histogram method is the simplest method to estimate a density.
- Suppose x is the random variable that we want to estimate its density using an i.i.d. sample with length n.
- The histogram estimate of the density is computed by splitting the range of x into equally spaced intervals and calculating the fraction of sample in each interval.
- The length of the bin is called **bandwidth** and it is denoted h.

Histogram

- We also need to adjust for the width of the bin so that the relative frequency can be compared across bins of different sizes.

- This implies that the area of the rectangle is $(freq/Nh) * h = freq/N$. Notice that this approximates the concept of probability derived from the density functions:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

  The area under the curve at a particular interval equals the probability.

Histogram

- We can define the intervals as:

$$B_k := [t_k, t_{k+1}) : t_k = t_0 + h_k, k \in \mathbb{Z}]$$

- And the histogram estimate is then defined as:

$$\hat{f}_H(x; t_0, h) := \frac{1}{nh} \sum_{i=1}^{n} \mathbf{1}\{X_i \in B_k : x \in B_k\}$$

- If we define the number of elements in a particular interval as v_k then

$$\hat{f}_H(x; t_0, h) = \frac{v_k}{nh}, \text{ if } x \in B_k \text{ for a certain } k \in \mathbb{Z}$$

Histograms. Stata aplication.

Stata

Kernel Estimation

The problem with histograms is that they are not very smooth, even with smalls h.

Another problem is that they depend on the choice of t_0.

The first improvement is the **moving histogram**:

- Instead of create intervals based on t_0, let's aggregate our data in intervals of the form $(x-h, x+h)$. We can approximate the density as:

$$\begin{aligned} f(x) &= F'(x) \\ &= \lim_{h\to 0^+} \frac{F(x+h)-F(x-h)}{2h} \\ &= \lim_{h\to 0^+} \frac{P[x-h<X<x+h]}{2h} \end{aligned}$$

Kernel Estimation

Then our density estimate is:

$$\hat{f}_N(x;h) = \frac{1}{2nh}\sum_{i=1}^{n} 1\{x-h < X_i < x+h\}$$

Which can be written as:

$$\hat{f}(x;h) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)$$

With $K(z) = \frac{1}{2}\mathbf{1}_{\{-1<z<1\}}$ which is the uniform density.

But why should we give the same weight to all the observations that belong the interval $(x-h, x+h)$? This is how the **kernel density** is born!

Kernel Density Estimation: Stata implementation

Stata

Confidence Interval Estimation

Confidence Interval Estimation

Confidence interval estimation of kernel densities has some complications. This is because:

- The kernel estimate is known to be biased, denoted $b(x)$.
- The problem of the bias arises in finite samples when creating confidence intervals.

$$f(x_0) \in \widehat{f}(x_0) - b(x_0) \pm 1.96 \times \sqrt{\widehat{var[f(\hat{x}_0)\,]}}$$

Confidence Interval Estimation

To perform inference we need an **asymptotically pivotal** statistic.

This is an statistic whose limit distribution does not depend on unknown parameters. In our case we use:

$$Z(x_0) \equiv \frac{\hat{f}(x_0) - f(x_0) - b(x_0)}{\sigma(x_0)} = \frac{\hat{f}(x_0) - E\{\hat{f}(x_0)\}}{\sigma(x_0)} \xrightarrow{d} N(0,1)$$

Where the statistic is known provided that we know the variance $\sigma^2(x_0)$. Unfortunately this is an unknown parameter and we therefore use its sample analog.

$$t(x_0) = \frac{\hat{f}(x_0) - E\{\hat{f}(x_0)\}}{s(x_0)}$$

Confidence Interval Estimation

- Under asymptotic theory we know that statistic will behave like standard normal.
- However, when the sample is finite we don't know its distribution and we can't get the critical values.
- To overcome this problem we usually approximate the distribution with its asymptotic distribution.
- The problem is that since the kernel estimate converges slower than other parametric estimates, the approximation errors gets amplified.
- To overcome this problem we can use resampling methods such as **jacknife** or **bootstrap** to estimate the distribution of $t(x_0)$.

Jacknife

Statistical resampling consists on the creation of new samples based on one observed sample.

They allow us to do statistical inference whenever this is not possible.

Given a sample of size n, **Jacknife** consists on creating different samples of size $n - 1$ by omitting one observation.

Bootrap Methods

There are several resampling options, the most known consists on treating the sample as the population, and create different new samples from the original sample with replacement.

It can be used for (among others):

- The estimate $\hat{\theta}$
- Standard errors $s_{\hat{\theta}}$
- T-statistic $t = (\hat{\theta} - \theta_0)/s_{\hat{\theta}}$, where θ_0 is the null hypothesis value
- The associated critical value or p-value for this statistic
- A confidence interval

Bootstrap Example

- Goal: Estimate variance of the sample mean $\bar{y}$ when y_i it is not known that $V[\bar{y}] = \sigma^2/N$.
- **Problem**: to compute the variance of the sample mean we need a sample of sample means.
- Bootstrap can implement this approach by viewing the sample as the population and drawing B bootstrap samples of size N with replacement from this population.
- Calculate the sample mean $\bar{y}_b$ for each bootstrap sample, and estimate $V[\bar{y}] = \frac{1}{B-1} \sum b = 1^B (\bar{y}_b - \bar{y})^2$.

Bootstrap Algorithm

- Given data $w_1, \ldots, w_N$
  - Draw a bootstrap sample of size N (see below)
  - Denote this new sample $w_1^*, \ldots, w_N^*$.
- Calculate an appropriate statistic using the bootstrap sample.
  - Examples include:
    (a) Estimate $\hat{\theta}$ of θ
    (b) Standard error $s_{\hat{\theta}}$ of estimate $\hat{\theta}$
    (c) t-statistic $t = (\hat{\theta} - \theta)/s_{\hat{\theta}}$ centered at $\hat{\theta}$
- Repeat steps 1-2 B independent times.
  - Gives B bootstrap replications of $\hat{\theta}_1, \ldots, \hat{\theta}_B$ or $t_1, \ldots, t_B$ or ...
- Use these B bootstrap replications to obtain a bootstrapped version of the statistic.

Bootstrap

We can apply the previous procedure to perform bootstrap on the statistic:

$$t^*(x_0) = \frac{\hat{f}^*(x_0) - \hat{f}(x_0)}{s^*(x_0)}$$

where $*$ denotes the bootstrap version.

The critical values (at level α) are the lower $\alpha/2$ and upper $\alpha/2$ quantiles of the ordered test statistics t^*. The **percentile-t method** $100(1-\alpha)$ percent confidence interval is:

$$\left(E\{\hat{f}(x_0)\} - t^*_{1-\alpha/2} \cdot s(x_0), ; E\{\hat{f}(x_0)\} + t^*_{\alpha/2} \cdot s(x_0)\right)$$

Bootstrap

Stata codes how to bootrsap

Confidence Interval Estimation

We have dealt with the first problem, which was about knowing the distribution of $t(x_0)$. What about the bias?

We can deal with this bias in different ways:

- Explicit bias removal.
- Undersmoothing.

Explicit bias removal

This is done by characterizing the analytical expression of the bias.

This has some problems since this involves simulating the second derivative of the estimated density.

$$b(x_0) = E[\hat{f}(x_0)] - f(x_0) = \frac{1}{2}h^2 f''(x_0) \int_{-\infty}^{\infty} z^2 K(z)\, dz$$

Undersmoothing

Remember that the optimal h^* minimizes the mean squared error, which is a combination of bias and variance.

What if we put more emphasis on minimizing the bias? We can get a kernel estimate whose bias converges faster to 0, therefore reducing the bias under a finite sample.

Since the bias converges to 0 as h converges to 0, we can choose a smaller h than h^* to draw the confidence intervals.

This is why we call this method undersmoothing. The confidence interval becomes:

$$(\hat{f}_{us}(x_0) - s_{us}(x_0) \times t_{us,1-\frac{\alpha}{2}}, \hat{f}_{us}(x_0) - s_{us}(x_0) \times t_{us,\frac{\alpha}{2}})$$

Important! The bias is still there, is just that it is smaller.