Section 1: LLN, CLT, Slutsky

Valentine Gilbert

September 10, 2021

Overview

Your TF

- G5 in Public Policy
- Fields are labor and public economics
- Research is on urban residential sorting
- Third year TFing Econ 2110 and 2115
- Dog dad

Figure: Lily

Section and General Advice

Section Goals:

- Build **intuition** around material covered in class
- **Review** concepts needed for problem sets
- Answer questions
- I think office hours are a better venue for discussing coding

Advice:

- Don't worry if the material feels difficult – it's supposed to
- Don't hesitate to **ask questions**! Use Slack, email me, come to office hours
- Review problem set solutions, even if you didn't lose points
- Review lecture notes

The Big Picture

In class we:

- Learned about the LLN, which says the sample mean is a consistent estimator of the expected value
- Learned about the CLT, which says sample mean is asymptotically normal
- Learned about Slutsky's Theorem

Why?

- The LLN, CLT, and Slutsky's Theorem are building block tools
- We'll use them to show when other estimators (e.g. regression coefficients) are consistent
- We'll use them to understand how much we learn from our estimates

Estimands, Estimators, and Estimates

An **estimand** is the thing you're trying to estimate

- In class, our estimands were $E[Y]$ and $var[Y]$
- Later in this class and in your own research, estimands will be causal effects

An **estimator** is a function that maps the observed data $Y_1, Y_2, ..., Y_N$ to a number (usually)

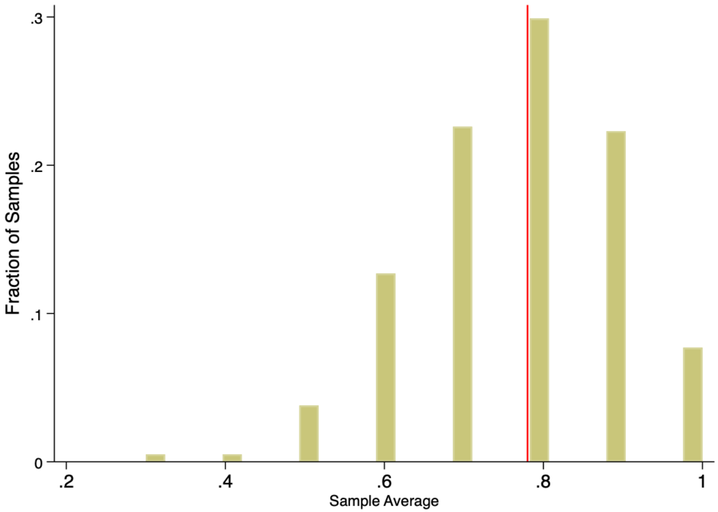
- E.g. $\bar{Y} = f(Y_1, Y_2, ..., Y_N) = \frac{1}{N}\sum_i Y_i$

An **estimate** is one realization of the estimator (using the data we actually observe)

$\Longrightarrow$ An estimate is a random variable! It has a distribution

- E.g. an estimate has a standard deviation, which we call its *standard error*

Sampling distribution of share of heads from 10 flips of a biased coin



Independence

Independence of random variables comes up often in this class. What does independence mean?

- Two random variables X and Y are *independent* if $f_{X|Y}(x|y) = f_X(x)$ and $f_{Y|X}(y|x) = f_Y(y)$

  $\implies$ Knowing $Y = y$ tells you nothing about the probability that $X = x$; knowing $X = x$ tells you nothing about the probability that $Y = y$
- An equivalent definition of independence is that the joint density is equal to the product of the marginal densities: $f_{XY}(x, y) = f_X(x)f_Y(y)$
- E.g. a coin toss and the roll of a die are independent events
- E.g. the number of heads after two coin flips and the outcome of the first coin flip are not independent

Question: Suppose you observe the i.i.d. data $Y_1, Y_2, ..., Y_N$ and calculate the sample mean, $\bar{Y}$. Are observations of $Y_i - \bar{Y}$ independent?

Bias and Consistency: Convergence in Probability

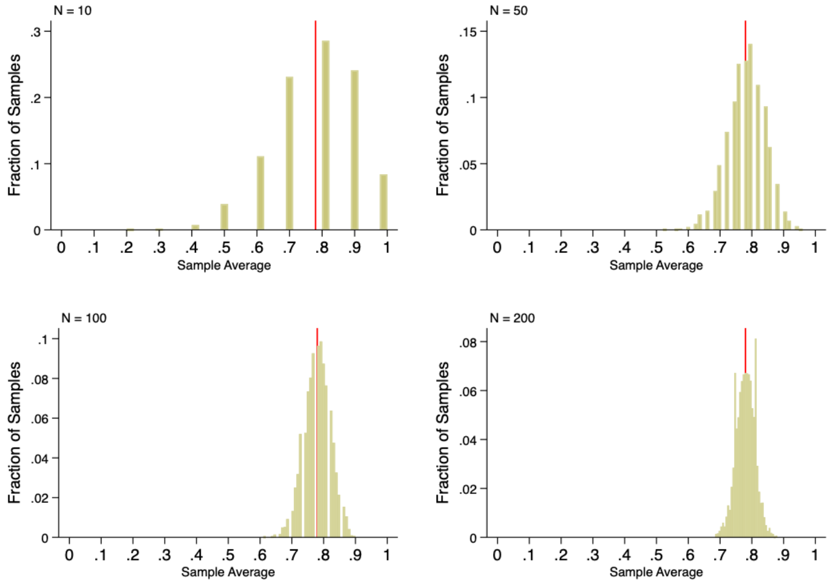
We're going to use **asymptotics** to understand different properties of our estimators. That is, we're going to ask what happens as our sample size gets large.

- Consider a sequence of sample statistics, $T_1, T_2, ..., T_N$, where the subscript indexes the sample size used to estimate the statistic
  - e.g. T_{10} is calculated using a sample of size 10
- T_N **converges in probability** to a constant c if and only if

$$\Pr(|T_N - c| \geq \delta) \to 0 \text{ as } N \to \infty \text{ for all } \delta > 0$$

- We can write this either as $T_N \xrightarrow{p} c$ or plim $T_N = c$

In words, a sample statistic converges in probability to a constant if you are less and less likely to calculate a sample statistic that is very far from the constant as you get more and more data. I.e., its sampling distribution becomes more and more concentrated around c.


N = 10
N = 50
N = 100
N = 200
Fraction of Samples
Sample Average
0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1

Bias and Consistency

We are now ready to define consistency and unbiasedness:

- T_N is a **consistent** estimator of the population parameter θ if $T_N \xrightarrow{p} \theta$
- T_N is an **unbiased** estimator of the population parameter if $E[T_N] = \theta$

Note that consistency is an asymptotic property whereas unbiasedness can hold in finite samples

- We saw in class that $\bar{Y}_N$ is an unbiased estimator of $E[Y]$ regardless of sample size

Bias and Consistency

It is possible for an estimator to be **consistent but biased** or **unbiased but inconsistent**

Consider $\hat{\mu} \equiv \bar{X}_N + \frac{1}{N}$

- $E[\hat{\mu}] = E[\bar{X}_N] + \frac{1}{N} = E[X] + \frac{1}{N} \neq E[X]$

  $\implies$ $\hat{\mu}$ is *biased* for any sample size
- But what happens as $N \to \infty$? The sample mean converges in probability to $E[X]$ and the term $\frac{1}{N} \to 0$

  $\implies$ $\hat{\mu}$ is *consistent*

Now consider $\tilde{\mu} \equiv X_1$, where X_1 is the first observation in the data

- $E[\tilde{\mu}] = E[X_1] = E[X]$

  $\implies$ $\tilde{\mu}$ is *unbiased*
- But the variance of our estimate doesn't decrease as $N \to \infty$

  $\implies$ $\tilde{\mu}$ is *inconsistent*

Jensen's Inequality: Definition

- **Jensen's Inequality** is a useful result that you'll need to apply on this week's homework
- It also appears in other areas of economics, such as choice under uncertainty and optimal taxation theory
- It states

$$E[h(Y_i)] \leq h(E[Y_i]) \text{ for concave } h$$

In words, the expectation of a concave function of a random variable is less than or equal to the function evaluated at the expectation of the random variable.

This inequality holds strictly for strictly concave h and is flipped for convex h.

Jensen's Inequality: Graphical Intuition

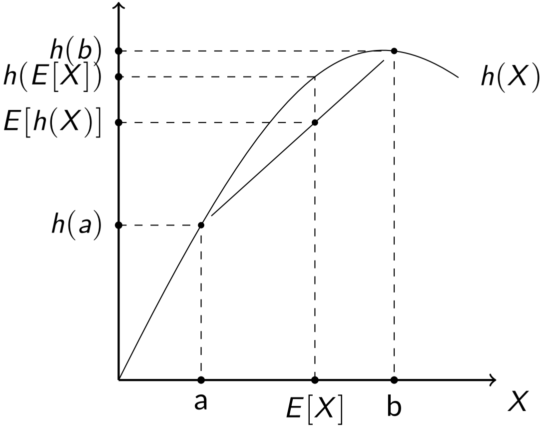


Figure: Jensen's Inequality for a binary random variable

The LLN

The **Law of Large Numbers** says that under certain conditions, the sample average is a consistent estimator of the the expected value. More formally,

- Suppose $Y_1, ..., Y_N$ are i.i.d., $E[Y_i] = a$, and $var(Y_i) < \infty$
- Then $\bar{Y}_N \xrightarrow{p} a$

This holds for more complicated looking random variables, too:

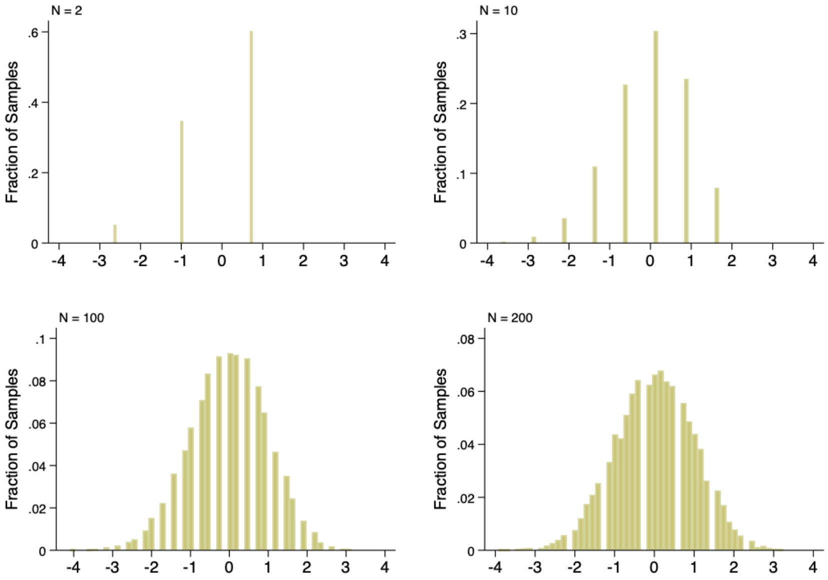
- Suppose X_i and Y_i are random variables and $W_i \equiv X_i Y_i$ is i.i.d. and meets our other conditions for the LLN
- Then $\frac{1}{N} \sum_i X_i Y_i \xrightarrow{p} E[X_i Y_i]$

The CLT

- The **Central Limit Theorem** tells us about the *shape* of the sample mean's distribution as the sample grows large
- Question: Why do we care about the shape of the distribution?
- The CLT says that if $Y_1, \ldots, Y_N$ are i.i.d. and if $\sigma_Y^2 < \infty$, then

$$\sqrt{N}(\bar{Y}_N - \mu)/\sigma \xrightarrow{d} N(0, 1)$$

- Question: Why do we have to multiply by $\sqrt{N}$?



N = 2
N = 10
N = 100
N = 200
Fraction of Samples
0
.2
.4
.6
.1
.3
.02
.04
.06
.08
-4
-3
-2
-1
0
1
2
3
4

Slutsky's Theorem

Slutsky's Theorem lets us apply the LLN and the CLT to functions of random variables. It has 4 parts:

1. If $T_N \xrightarrow{p} c$ and $h(T_N)$ is a continuous function, then $h(T_N) \xrightarrow{p} h(c)$. Another way of writing this is $\text{plim}\ h(T_N) = h(\text{plim}\ T_N)$.
2. If $V_N \xrightarrow{p} c_1$, $W_N \xrightarrow{p} c_2$, and $h(V_N, W_N)$ is a continuous function, then $h(V_N, W_N) \xrightarrow{p} h(c_1, c_2)$. Another way of writing this is $\text{plim}\ h(V_N, W_N) = h(\text{plim}\ V_N, \text{plim}\ W_N)$.
3. If $V_N \xrightarrow{p} c$ and W_N has a limiting distribution, then the limiting distribution of $V_N + W_N$ is equal to the limiting distribution of $c + W_N$
4. If $V_N \xrightarrow{p} c$ and W_N has a limiting distribution, then the limiting distribution of $V_N W_N$ is equal to the limiting distribution of cW_N

Application: Consistency of the Sample Variance

Let's review how we can apply the LLN and Slutsky to show that the sample variance $\widehat{var(Y_i)} = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$ is consistent. With some algebra, we can rewrite this as

$$\frac{1}{N}\sum_{i=1}^{N}\left((Y_i - E[Y_i])^2\right) - (\bar{Y} - E[Y_i])^2$$

Our strategy for proving consistency is to:

1. Show that the second term converges in probability to 0
2. Show that the first term converges in probability to the population variance
3. Combine those results to show that the sample variance converges in probability to the population variance

Application: Consistency of Sample Variance

$$\frac{1}{N}\sum_{i=1}^{N}\left((Y_i - E[Y_i])^2\right) - (\bar{Y} - E[Y_i])^2$$

Let's begin with the second term: $(\bar{Y} - E[Y_i])^2$

- By the LLN, $\bar{Y} \xrightarrow{p} E[Y_i]$
- Applying Slutsky 1 to the function $f(a) = (a - E[Y_i])^2$, we know that $f(\bar{Y}) \xrightarrow{p} f(E[Y_i])$
- That is, $(\bar{Y} - E[Y_i])^2 \xrightarrow{p} (E[Y_i] - E[Y_i])^2 = 0$

$\implies$ This term converges in probability to 0!

Application: Consistency of Sample Variance

$$\frac{1}{N}\sum_{i=1}^{N}\left((Y_i - E[Y_i])^2\right) - (\bar{Y} - E[Y_i])^2$$

Now let's turn to the first term: $\frac{1}{N}\sum_{i=1}^{N}\left((Y_i - E[Y_i])^2\right)$

- Define $W_i \equiv (Y_i - E[Y_i])^2$
- The random variables $W_1, \ldots, W_N$ are i.i.d., so we can apply the LLN

$\implies \quad \frac{1}{N}\sum_{i=1}^{N} W_i \xrightarrow{p} E[W_i] = E[(Y_i - E[Y_i])^2] = var(Y_i).$

Application: Consistency of Sample Variance

$$\frac{1}{N}\sum_{i=1}^{N}\left((Y_i - E[Y_i])^2\right) - (\bar{Y} - E[Y_i])^2$$

Now we just have to use Slutsky 2 to combine these results:

- Let $g(a, b) = a - b$
- By Slutsky 2, the probability limit of our entire expression is just the difference between the probability limits of our two terms

$\implies \widehat{var(Y_i)} \equiv \frac{1}{N}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$ is a consistent estimator of σ_Y^2!

Question: Why couldn't we have defined $V_i \equiv (Y_i - \bar{Y})^2$ and applied the law of large numbers directly to the sample variance?