Econometrics I

Lecture 2: Asymptotic Statistics

Chris Conlon
NYU Stern
Fall 2025

Roadmap

1. Sample and population means
2. Convergence in probability
3. Law of Large Numbers
4. Convergence in distribution
5. Central Limit Theorem

Convergence

- Consider a sequence of numbers $\{X_i\}_{i=1}^n = (X_1, X_2, \ldots, X_n)$.
  - Often people write $X_n = \{X_i\}_{i=1}^n$ but I hate this.
- We can ask what happens to some function $g(\{X_i\}_{i=1}^n)$ as $n \to \infty$.
  - Simple Example: $X_i = \frac{1}{2^i}$ and $g(\cdot)$ is the summation so that $\lim_{n\to\infty} \sum_{i=1}^n \frac{1}{2^i} = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots \to 1$.
- Sequences of random variables "approach" something too (other random variables)
- But what does approaching mean?

Convergence

Let's define $S_n = g(\{X_i\}_{i=1}^n)$ as some statistic of the sequence of random variables $\{X_i\}_{i=1}^n$.
For example:

- Summation: $S_n = \sum_{i=1}^n X_i$; Sample Average: $S_n = \frac{1}{n}\sum_{i=1}^n X_i$;
- Largest Element: $S_n = \max X_i$
- Variance: $Var(X_1, X_2, \ldots, X_n)$.
- Something complicated: $X_1 - 2X_2 + 3X_3 - 4X_4 \ldots$.

But our statistic could just be the identity function $S_n = \{X_i\}_{i=1}^n$ or transform $S_n = \{aX_i + b\}_{i=1}^n$

Convergence in Probability

We say that the statistic S_n converges in probability to S which we write as $S_n \xrightarrow{p} S$ if

$$\lim_{n \to 0} \mathbb{P}\left(\|S_n - S\| > \epsilon\right) \to 0$$

- This means that the probability that random variables S_n and S are far apart will not be large as $n \to \infty$.
- Continuous mapping theorem says if: $S_n \xrightarrow{p} S$ then $g(S_n) \xrightarrow{p} g(S)$ if $g(\cdot)$ is a continuous function.
- Example: X_i is i.i.d $\mathcal{N}(\mu, \sigma^2)$ what does $S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ converge in probability to?
- Example: $X_n = \frac{1}{n}$ with probability $1 - \frac{1}{n}$ and n with probability $\frac{1}{n}$. What does this converge in probability to?

Convergence Almost Surely

We say that the statistic S_n converges almost surely to S which we write as $S_n \overset{a.s.}{\to} S$ if

$$\mathbb{P}\left(\lim_{n\to\infty} S_n = S\right) = 1$$

- This means that the probability that random variables S_n is equal to S approaches one as as $n \to \infty$.
- This implies convergence in probability (but not *vice versa*). It is stronger convergence

Convergence In Distribution

We say that the statistic X_n converges in distribution to S which we write as $X_n \xrightarrow{d} X$ if

$$\lim_{n\to\infty} F_n(x) = F(x) \text{ at all points } x \text{ where } F \text{ is continuous.}$$

- This means that as $n \to \infty$ their c.d.f's agree (but not necessarily their p.d.f.'s)
- Continuous mapping theorem says if: $X_n \xrightarrow{d} X$ then $g(X_n) \xrightarrow{d} g(X)$ if $g(\cdot)$ is a continuous function.
- This is a weaker sense of convergence and is implied by convergence in probability (but not *vice versa*).
- A common goal is to show that some statistic of the sequence $\{X_I\}_{i=1}^n$ is distributed normally.

Convergence: comments

- For a constant k,

$$X_N \overset{a.s.}{\longrightarrow} k \qquad \text{implies} \qquad X_N \overset{p}{\longrightarrow} k \qquad \text{iff} \qquad X_N \overset{d}{\longrightarrow} k$$

- Note that
  - convergence in probability usually comes up in the context of X_N converging to a **fixed value**;
  - convergence in distribution usually comes up in the context of X_N converging to a **random variable**.

The Big Theorems

Markov Inequality

Let X be a non-negative random variable with finite expectation $\mathbb{E}[X]$ then for any $a > 0$

$$\mathbb{P}(X \geqslant a) \leqslant \frac{\mathbb{E}[X]}{a}$$

This basically says "if $\mathbb{E}[X]$ is small, the it is unlikely that X is large".

Proof: Define Y so that $X \geqslant Y$:

$$Y = \begin{cases} a & X \geqslant a \\ 0 & X < a \end{cases}$$

Note that: $\mathbb{E}[X] \geqslant \mathbb{E}[Y] = a \cdot \mathbb{P}\{X \geqslant a\}$. and divide by a.

Markov Inequality

Let X be a non-negative random variable with finite expectation $\mathbb{E}[X]$ then for any $a > 0$

$$\mathbb{P}(X \geqslant a) \leqslant \frac{\mathbb{E}[X]}{a}$$

This basically says "if $\mathbb{E}[X]$ is small, the it is unlikely that X is large".

Proof: Define Y so that $X \geqslant Y$:

$$Y = \begin{cases} a & X \geqslant a \\ 0 & X < a \end{cases}$$

Note that: $\mathbb{E}[X] \geqslant \mathbb{E}[Y] = a \cdot \mathbb{P}\{X \geqslant a\}$. and divide by a.

Chebyshev Inequality

Let X be a non-negative random variable with finite expectation $\mathbb{E}[X]$ and variance $E[(X-\mu)^2] = \sigma^2$.

$$P\{|X-\mu| \geqslant k\} \leqslant \frac{\sigma^2}{k^2}$$

This basically says "if $Var(X)$ is small, the it is unlikely that X is too far from the mean".

Proof: Define $Z = (X-\mu)^2$ as our non-negative random variable.

$$P\{|X-\mu| \geqslant k\} = P\left\{(X-\mu)^2 \geqslant k^2\right\}$$

Now apply Markov inequality to Z where $a = k^2$.

Chebyshev Inequality

Let X be a non-negative random variable with finite expectation $\mathbb{E}[X]$ and variance $E[(X-\mu)^2] = \sigma^2$.

$$P\{|X-\mu| \geqslant k\} \leqslant \frac{\sigma^2}{k^2}$$

This basically says "if $Var(X)$ is small, the it is unlikely that X is too far from the mean".

Proof: Define $Z = (X-\mu)^2$ as our non-negative random variable.

$$P\{|X-\mu| \geqslant k\} = P\left\{(X-\mu)^2 \geqslant k^2\right\}$$

Now apply Markov inequality to Z where $a = k^2$.

The Law of Large Numbers

Weak Law Of Large Numbers
If X_i for $i = 1, \ldots, N$ are independent and identically distributed with finite mean μ and variance σ^2, then as $N \to \infty$,

$$\bar{X}_N \xrightarrow{p} E[X]$$

The Law of Large Numbers says the sample average converges in probability to its expectation.

- If we flip a fair coin enough times $\frac{1}{N}\sum_{i=1}^{n} X_i \xrightarrow{p} \frac{1}{2}$ heads.
- If we roll fair 6-sided die enough times $\frac{1}{N}\sum_{i=1}^{n} X_i \xrightarrow{p} 3.5$.
- If we randomly draw enough $\mathcal{N}(\mu, \sigma^2)$ variables on our computer $\frac{1}{N}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu$.
- What if instead we asked $\frac{1}{N}\sum_{i=1}^{n} (X_i - \mu)^2 \xrightarrow{p} ?$.

Proof: Law of Large Numbers

$$S_n = \frac{1}{n}(X_1 + X_2 + X_3 + \cdots + X_n).$$

Easy to see that $\mathbb{E}[S_n] = \mu$, and that

$$\begin{aligned} Var(S_n) &= Var\left(\frac{1}{N}(X_1 + X_2 + \cdots X_n)\right) \\ &= \frac{1}{n^2} Var\left(X_1 + X_2 + \cdots X_n\right) = \frac{1}{n}\sum_{i=1}^{n} Var(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \end{aligned}$$

Now we apply Chebyshev with $k = \epsilon$:

$$\mathbb{P}\left(|S_n - \mu| \geqslant \varepsilon\right) \leqslant \frac{\sigma^2}{n \cdot \varepsilon^2}$$

Which goes to zero as $n \to \infty$.

LLN generalization: intuition

- We argued above that the variance of the sample mean converges to zero; Chebyshev's inequality allows us to make the leap from this fact to convergence in probability of sample means to population means.
- Notice that we didn't need to know the distribution of X_i. For a generic distribution (with finite variance), we would not necessarily know the distribution of the sample mean, but it's often the case that the sample mean of i.i.d. variables will have expectation μ, and its variance will be proportional to $1/N$.

- For i.i.d. sequences (with finite first and second moments), sample means are unbiased estimators of population means, and their variance converges to zero.
- $\hat{\theta}$ is an unbiased estimator of θ if $\mathbb{E}\left[\hat{\theta}\right] = \theta$.

Unbiasedness and Consistency

- As mentioned above, $\hat{\theta}$ is an **unbiased** estimator of θ if $E\left[\hat{\theta}\right] = \theta$.
- $\hat{\theta}$ is a **consistent** estimator of θ if $\hat{\theta} \xrightarrow{p} \theta$ as $N \to \infty$
- While closely related, note that neither implies the other. Examples?

Central Limit Theorem

- If X_i are i.i.d. draws with finite mean and variance, then as $N \rightarrow \infty$,

$$\sqrt{N}\left(\bar{X}_N - \mu\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right)$$

  where $\mu = E\left(X_i\right)$ and $\sigma^2 = Var\left(X_i\right)$.

- Note that the CLT tells us something very specific about the behavior of the variance of $\bar{X}_N$ as N gets large: **root-N convergence**

Moment generating functions revisited

Levy's continuity theorem (MGF version)
Let X_n have MGF $M_n(t)$. If $M_n(t)$ converges pointwise to $M(t)$, the MGF of X, then

$$X_n \xrightarrow{d} X$$

- Recall: when two variables have same MGF, they have the same distribution
- Idea for CLT proof: show that MGF of X_N becomes the normal MGF

I am not going to prove the CLT (sorry!).

LLN and CLT Illustration

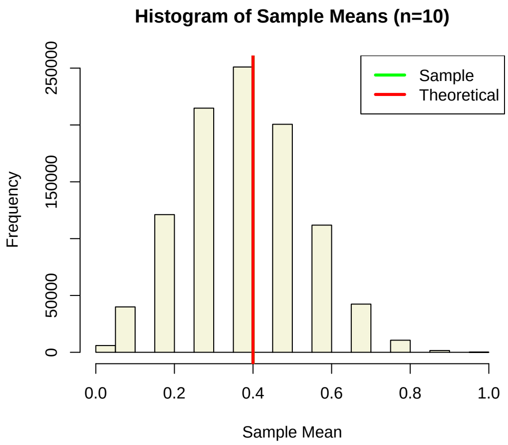
- Simulate $X_i \sim \text{Bernoulli}(p)$ for $i = 1, 2, \ldots, N$.
- Compute $\bar{X}_N = N^{-1} \sum_{i=1}^{N} X_i$. Note that
  - $\bar{X}_N$ is N^{-1} times a random variable that is $\sim \text{Binomial}(N, p)$.
  - $E\left[\bar{X}_N\right] = p$.
- Repeat above $S = 1000$ times. Let $\bar{X}_{N,s}$ denote the sample mean from the sth simulation
- I'm going to do this and plot the empirical distribution (histograms) of $\bar{X}_{N,s}$ across s

LLN Illustration

```
clt_function <- function(n_sample,p){
  # create an empty vector for the means binomial samples
  clt <- NULL
  n <- n_sample   # sample size for each simulation
  # take the mean ofsamples of the uniform distribution. repeat 1,000 times
  for (i in 1:1000) {
    clt <-  c(clt, rbinom(1000,n_sample,p)/n_sample)
  }
  theoretical_mean <- p
  hist(clt, xlim=c(0, 1), xlab='Sample␣Mean', main=paste0("Histogram␣of␣Sample␣Means␣(n=", eval(n_
       sample),")"), col='beige')
  abline(v=mean(clt), lwd=3, col='green')
  abline(v=theoretical_mean, lwd=3, col='red')
  legend(c("Sample", "Theoretical"),x='topright', lwd=c(3,3), col=c('green', 'red'))
}
```

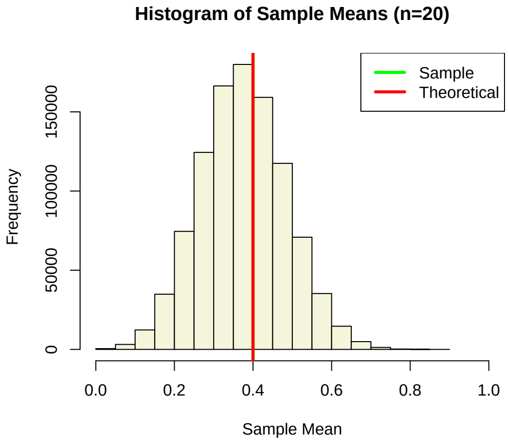
LLN Illustration

```
clt_function(10,.4)
```



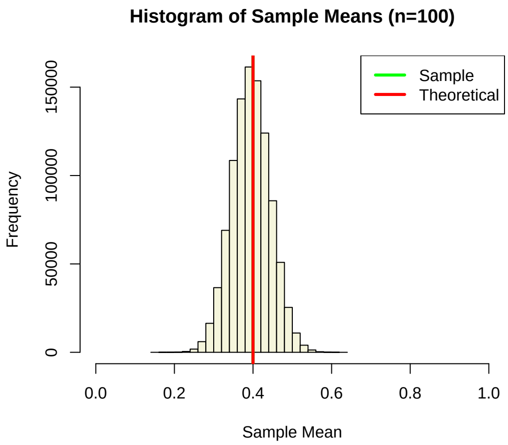
LLN Illustration

```
clt_function(20,.4)
```



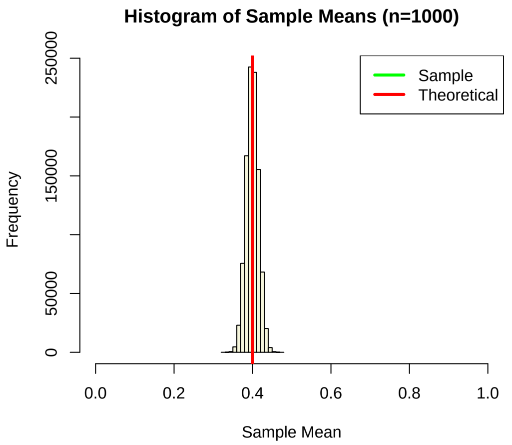
LLN Illustration

```
clt_function(100,.4)
```



LLN Illustration

```
clt_function(1000,.4)
```

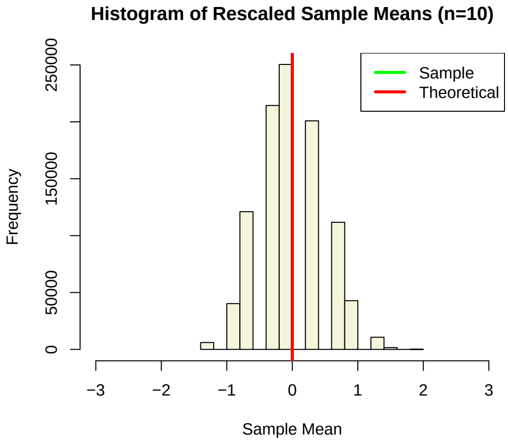


CLT Illustration

```
clt_function_scaled <- function(n_sample,p){
  # create an empty vector for the means binomial samples
  clt <- NULL
  n <- n_sample   # sample size for each simulation
  # take the mean ofsamples of the uniform distribution. repeat 1,000 times
  for (i in 1:1000) {
    clt <-  c(clt, (rbinom(1000,n_sample,p)/n_sample-p)*sqrt(n_sample))
  }
  theoretical_mean <- 0
  hist(clt, xlim=c(-3, 3), xlab='Sample␣Mean', main=paste0("Histogram␣of␣Standardized␣Sample␣Means␣(n
      =", eval(n_sample),")"), col='beige')
  abline(v=mean(clt), lwd=3, col='green')
  abline(v=theoretical_mean, lwd=3, col='red')
  legend(c("Sample", "Theoretical"),x='topright', lwd=c(3,3), col=c('green', 'red'))
}
```

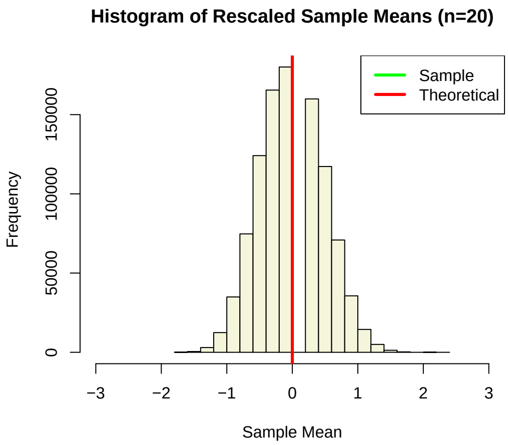
CLT Illustration

```
clt_function_scaled(10,.4)
```



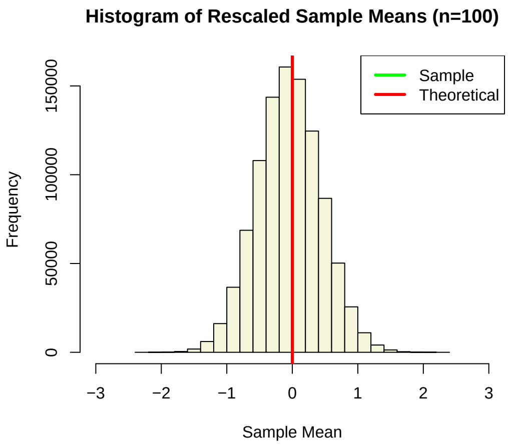
CLT Illustration

```
clt_function_scaled(20,.4)
```



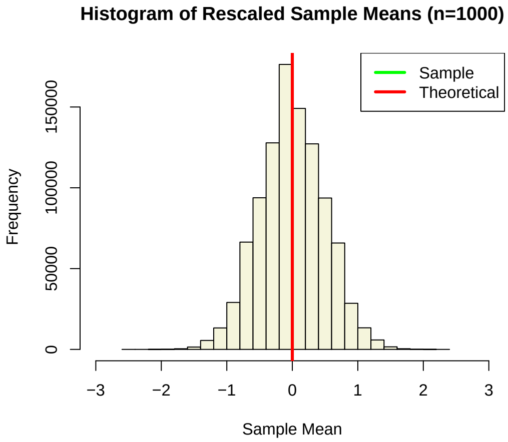
CLT Illustration

```
clt_function_scaled(100,.4)
```



CLT Illustration

```
clt_function_scaled(1000,.4)
```



CLT interpretation

$$\bar{X}_N \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{N}\right)$$

- This is **not strictly correct**: the CLT just tells us that $\bar{X}_N$ will be approximately distributed $\mathcal{N}\left(\mu, \frac{\sigma^2}{N}\right)$ as N gets large, but in any finite sample the distribution of $\bar{X}_N$
  - might not be centered around μ
  - might not have variance σ^2/N
  - might not be normal
- for i.i.d. normal case, the above statement actually is strictly correct.
- In empirical economics, we generally use this approximation, but it's important to keep in mind that it's not always a good approximation, particularly for small sample sizes when the underlying distribution of the random variable is not itself normal.

CLT extensions

- Like the LLN, there are versions of the CLT that get away from both the **identical** and **independent** aspects of the version we presented. But we *do* need related conditions. See: **ergodicity**, **stationarity**, **weak dependence**.

Finite sample distributions

- It is typically difficult to say much about the distributions of statistics in **finite samples**.
- Exception: when the underlying random variables are normally distributed. Note that above, it was trivial to derive the distribution of $\bar{X}_N$ when X_i was i.i.d. normal.
- **Monte Carlo studies** (simulation studies) are typically used to explore the finite-sample properties of estimators.
- We'll talk (briefly) later about **bootstrapped standard errors**, which can have appealing finite sample properties.

Multivariate extensions

- Let $\{\mathbf{X}_i\}$ for $i = 1, 2, \ldots, \infty$ be an i.i.d. sequence of vectors of random variables with finite first and second moments.
  - Note that here we're asking for the $\mathbf{X}_i$ vector to be i.i.d. across i. Different elements of $\mathbf{X}_i$ need not be identical or independent.
- The extension of the LLN to **multivariate** random variables is trivial. We can write

$$\bar{\mathbf{X}}_N \xrightarrow{p} E\left(\mathbf{X}_i\right)$$

- The multivariate CLT holds that

$$\sqrt{N}\left(\bar{\mathbf{X}}_N - E\left(\mathbf{X}_i\right)\right) \xrightarrow{d} \mathcal{N}\left(0, \Sigma\right),$$

where $\Sigma = E\left[\left(\mathbf{X}_i - E\left(\mathbf{X}_i\right)\right)\left(\mathbf{X}_i - E\left(\mathbf{X}_i\right)\right)'\right]$

Using the CLT

$$\sqrt{N}\left(\bar{X}_N - \mu\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right)$$

- We typically use the CLT to estimate the (approximate) distribution of estimators.
- This means that we need an estimate of σ^2, and then σ^2/N estimates the variance of $\bar{X}_N$

Estimating variance

- Consider the usual setup with X_i i.i.d. for $i = 1, 2, \ldots, \infty$.
- Let's show that

$$E\left[\left(X_i - \bar{X}_N\right)^2\right] = \sigma^2 \left(\frac{N-1}{N}\right)$$

- It follows that $E\left[s_N^2\right] = \sigma^2$, where

$$s_N^2 \equiv \frac{1}{N-1} \sum_{i=1}^{N} \left(X_i - \bar{X}_N\right)^2$$

- But also notice that for large N, there is little difference between this estimator and one that divides by N.

Standard errors

$$s_N^2 \equiv \frac{1}{N-1} \sum_{i=1}^{N} \left(X_i - \bar{X}_N\right)^2$$

- We can estimate $Var\left(\bar{X}_N\right) = \frac{\sigma^2}{N}$ with $\frac{s^2}{N}$
- A **standard error** refers to our estimate of $\sqrt{Var\left(\bar{X}_N\right)}$:

$$\sqrt{\frac{s^2}{N}}$$

- Given the **asymptotic normality** of $\bar{X}_N$, these standard errors allow us to make statements about the (approximate) CDF of $\bar{X}_N$.
- Note that *standard errors* refer to an estimate of the *standard deviation* of an estimator.