L.3: Law of Large Numbers

Econometrics 1: ver. 2024 Fall Semester

Naoki Awaya

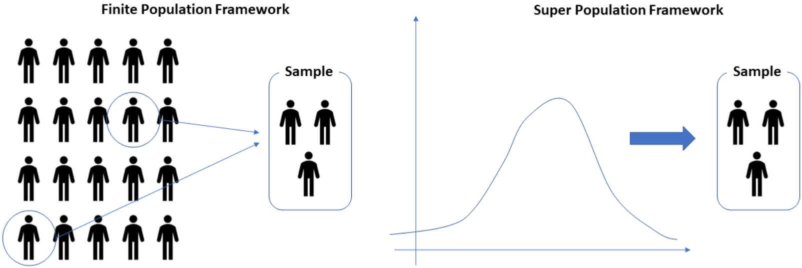
Population and Sample

Statistical Inference

- The entire set of objects we are interested in analyzing is called the population.
- Since the size of a population is often huge (it could be infinite), we use a small subset of the population, namely the sample, for statistical analysis.
    - Sampling: obtain a finite sample from the population of interest.
- Statistical inference: from the results on the sample, we can infer the population.

Statistical Inference

- There are two types of population concept:
  - finite-population = the population is a finite set (e.g., countries, prefectures, ...)
  - super-population = all observations following a population distribution (treat the underlying data dist as the target population itself)
- The observed finite-population can be regarded as a "sample" from a super-population.



Statistical Inference

- Throughout the lectures, we adopt the super-population framework (because of its analytical convenience).
- Statistical inference under a finite-population framework will be discussed in Econometrics 2.

The Law of Large Numbers: Introduction

The Law of Large Numbers: Introduction

LLN: Informal statement

Suppose that we have data $\{X_1, \dots, X_n\}$ of sample size n randomly drawn from the same population. Let

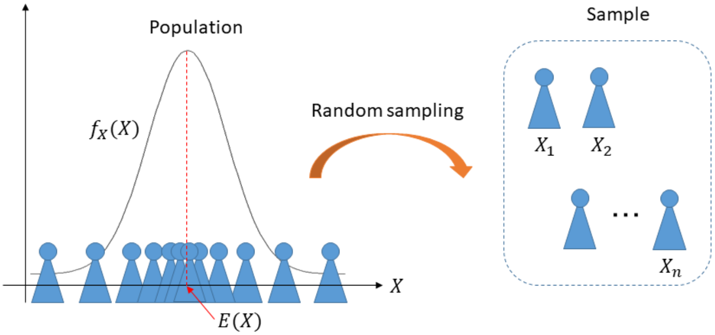
$$\text{Population mean of } X\text{: } \mu = \mathbb{E}(X)$$

$$\text{Sample average of } X\text{: } \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Then, as n increases to infinity, $\overline{X}_n$ approaches to μ.

When one considers a finite-population framework, n is fixed and cannot tend to infinity.

The Law of Large Numbers: Introduction



Population mean | Sample average

If n is sufficiently large. . . $E(X) \approx \frac{1}{n}\sum_{i=1}^{n} X_i$

Dice Roll Simulation

A simulation of dice rolling with **R**

- For example, if you want to simulate 10 dice rolls, execute the following code:

```
> library(extraDistr)
> Dice10 <- rdunif(10, 1, 6) # Draw from Unif{1,2, ..., 6}
> Dice10
```

```
##  [1] 3 4 2 4 6 2 2 1 6 3
```

```
> mean(Dice10)
```

```
## [1] 3.3
```

NOTE: Your results may be different from mine. To fix the simulation result, fix the "random seed" before generating random numbers.[1]

[1] You can use the `set.seed()` function.

Dice Roll Simulation

- Similarly, create `Dice20000`:

```
> Dice20000 <- rdunif(20000, 1, 6)
> mean(Dice20000)
```

```
## [1] 3.5001
```

- The sample average of `Dice20000` is roughly equal to 3.5.
- Recall that the expected value of a dice roll is $\mathbb{E}(\text{dice roll}) = 3.5$.
- That is, the sample average with $n = 20000$ is closer to 3.5 than the case with $n = 10$.
- This fact is known as the law of large numbers — *as the sample size increases, the sample average converges to its population mean*.

Dice Roll Simulation

- It would be helpful to visually demonstrate how the sample average converges to its mean.
- To this end, we can create a function whose input is the sample size and the output is the corresponding sample average.

```
>   Dice <- function(n){
+     Dicerolls <- rdunif(n, 1, 6)
+     mean(Dicerolls)
+     }
```

Dice Roll Simulation

- Then, if you specify any number for n, the function `Dice()` returns the sample average of dice rolls over n trials.

```
>    Dice(10)
```

```
## [1] 3.6
```

```
>    Dice(100)
```

```
## [1] 3.68
```

```
>    Dice(1000)
```

```
## [1] 3.493
```

Dice Roll Simulation

- Next, using the above created function `Dice()`, we repeat the dice rolling experiment for different n's: $n = 1, \ldots, N$.
- Such calculation can be performed using the for-loop command:

```
for(i in sequence)  statement
```

- Here, "`sequence`" is a vector, where `i` takes on each of its value during the loop.
- In each iteration, "`statement`" is executed.
- The iteration stops when `i` reaches to the final element of "`sequence`".

Dice Roll Simulation

- First, set any large number for N, and create a blank vector of length N (vector of zeros).

```
>   N <- 1000
>   R <- numeric(N) # N×1 vector or zeros
```

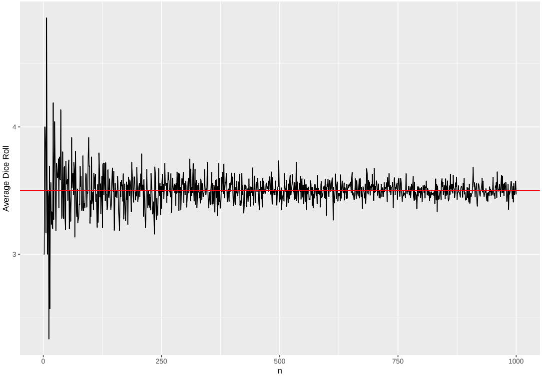
- For-loop iteration:

```
>   for(i in 1:N) R[i] <- Dice(i)
```

- Then, the i-th element of `R` contains the average value of i dice rolls.

Dice Roll Simulation

```
>   library(tidyverse)
>   data <- data.frame(n = 1:N, mean = R)
>   ggplot(data, aes(x = n, y = mean)) +
+     geom_line() + ylab("Average Dice Roll") +
+     geom_abline(intercept = 3.5, slope = 0, color = "red")
```



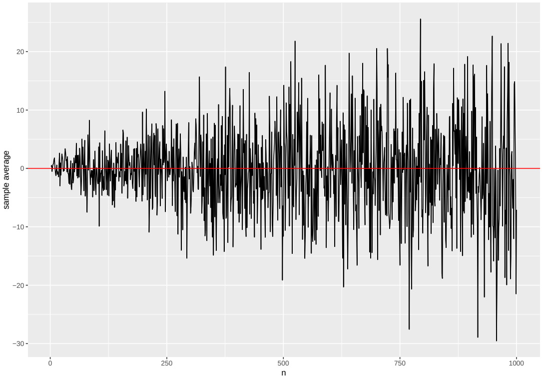
LLN is not always true

- In the following example, $\mathbb{E}[X_i] = 0$ for all i.

```
>   sim <- function(n){
+     X <- runif(1,-1,1)
+     for(i in 2:n) X <- c(X, X[i-1] + runif(1,-1,1))
+     mean(X)
+   }
>   N <- 1000
>   R <- numeric(N-1)
>   for(i in 2:N) R[i-1] <- sim(i)
```

LLN is not always true

```
>   data <- data.frame(n = 2:N, mean = R)
>   ggplot(data, aes(x = n, y = mean)) +
+     geom_line() + ylab("sample average") +
+     geom_abline(intercept = 0, slope = 0, color = "red")
```



The Law of Large Numbers: Introduction

- In what sense does the sample average converges to the population mean mathematically?
  $\Rightarrow$ Convergence in probability
- Under what conditions does LLN hold true?
  $\Rightarrow$ An important condition is that the data are drawn "independently" from the same population.[a]

[a] Independence is not a necessary condition, but it can greatly simplify the statistical analysis.

Convergence in probability

Convergence of real numbers

- Let $(a_n)_{n=1}^{\infty}$ denote a sequence of numbers: $(a_n)_{n=1}^{\infty} = \{a_1,\ a_2, \ldots\}$
- Example (i):

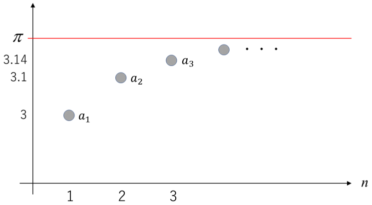
$$a_1 = 1,\ a_2 = 1.4,\ a_3 = 1.41\ \cdots a_n = 1.4142\ldots\cdots$$

a_n converges to $\alpha = \sqrt{2}$ as n increases to infinity.

- Example (ii):

$$a_1 = 3,\ a_2 = 3.1,\ a_3 = 3.14\ \cdots a_n = 3.1415\ldots\cdots$$

a_n converges to $\alpha = \pi$ as n increases to infinity.



Convergence of real numbers

- The number α is called the limit of $(a_n)_{n=1}^{\infty}$, and we write

$$\lim_{n\to\infty} a_n = \alpha \quad \text{or} \quad a_n \to \alpha \ (n \to \infty) \ \text{ in short}$$

Further, these are equivalent to

$$\lim_{n\to\infty} |a_n - \alpha| = 0 \quad \text{and} \quad |a_n - \alpha| \to 0 \ (n \to \infty).$$

- Note: a sequence does not always have a limit:

$$a_1 = -1,\ a_2 = 1,\ a_3 = -1, \cdots a_n = (-1)^n, \cdots$$

This sequence $(a_n)_{n=1}^{\infty}$ alternates between 1 and -1 and is not a convergent.

Convergence in probability

- Now, suppose $(v_n)_{n=1}^{\infty}$ is a sequence of "random variables".
- Then, since whether the sequence $(v_n)_{n=1}^{\infty}$ converges to α is random, we consider its probability.

Convergence in probability

A sequence of random variables $(v_n)_{n=1}^{\infty}$ is said to converge in probability to α if

$$\lim_{n\to\infty} \Pr\left(|v_n - \alpha| < \epsilon\right) = 1$$

for <u>any</u> positive $\epsilon > 0$.

Interpretation: v_n converges to α with probability tending to one as n increases.

Convergence in probability

- The most important point in this definition is that ϵ can be chosen arbitrarily small, say $\epsilon = 0.001$, $\epsilon = 0.00\ldots001$, or even smaller.
- However small ϵ is, we can observe $\{|v_n - \alpha| < \epsilon\}$ with almost 100% probability if n is sufficiently large.
- When $(v_n)_{n=1}^{\infty}$ converges to α in probability, we write

$$\text{plim}_{n\to\infty} v_n = \alpha \quad \text{or} \quad v_n \xrightarrow{p} \alpha \ (n \to \infty) \ \text{ in short}$$

  Here, plim is read as "probability limit".
- Note the following equivalence:

$$\lim_{n\to\infty} \Pr(|v_n - \alpha| < \epsilon) = 1 \iff \lim_{n\to\infty} \Pr(|v_n - \alpha| \geq \epsilon) = 0$$

Convergence in probability

LLN: Formal statement

Suppose we have data $\{X_1, \dots, X_n\}$ of sample size n independently drawn from the same population. Let

$$\text{Population mean of } X\text{: } \mu = \mathbb{E}(X)$$

$$\text{Sample average of } X\text{: } \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Then, we have

$$\overline{X}_n \xrightarrow{p} \mu \ (n \to \infty)$$

This result is known as the "weak" law of large numbers.

Strong LLN: $\Pr(\lim |\overline{X}_n - \mu| = 0) = 1.$

Proof of Weak LLN

Proof of Weak LLN

Mean and variance of sample mean

Suppose that we have data $\{X_1, \ldots, X_n\}$ of sample size n randomly drawn from the same population. Let $\mu = \mathbb{E}(X)$ and $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then,

$$\mathbb{E}\left(\overline{X}_n\right) = \mu, \quad \mathbb{V}\left(\overline{X}_n\right) = \frac{\mathbb{V}(X)}{n}$$

Proof.

- By assumption, X_i's are drawn from the same distribution:

$$\mathbb{E}(X_1) = \mathbb{E}(X_2) = \cdots = \mathbb{E}(X_n) = \mu$$

- Using the linearity of expectation, we have

$$\mathbb{E}\left(\overline{X}_n\right) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\underbrace{\sum_{i=1}^{n}\mathbb{E}(X_i)}_{=n\mu} = \mu$$

Proof of Weak LLN

- Similarly, since X_i's are drawn from the same distribution,

$$\mathbb{V}(X_1) = \mathbb{V}(X_2) = \cdots = \mathbb{V}(X_n) = \mathbb{V}(X)$$

- Note that, for any constant a,

$$\begin{aligned}\mathbb{V}(aX) &= \mathbb{E}(a^2X^2) - \mathbb{E}(aX)^2 \\ &= a^2\left[\mathbb{E}(X^2) - \mathbb{E}(X)^2\right] = a^2\mathbb{V}(X)\end{aligned}$$

- Thus, $\mathbb{V}(\overline{X}_n) = \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\mathbb{V}\left(\sum_{i=1}^{n} X_i\right)$.

Proof of Weak LLN

- Further, when X and Y are independent, $\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y)$.
- Repeatedly applying this result, we have

$$\begin{aligned}
\mathbb{V}\left(\sum_{i=1}^{n} X_i\right) &= \mathbb{V}(X_1) + \mathbb{V}\left(\sum_{i=2}^{n} X_i\right) \\
&= \mathbb{V}(X_1) + \mathbb{V}(X_2) + \mathbb{V}\left(\sum_{i=3}^{n} X_i\right) = \cdots = n\mathbb{V}(X).
\end{aligned}$$

- Thus, $\mathbb{V}(\overline{X}_n) = \frac{\mathbb{V}(X)}{n}$. ■

Proof of Weak LLN

Markov's inequality

Let X be a non-negative random variable, and $\epsilon > 0$ be a positive constant. Then,

$$\Pr\left(X \geq \epsilon\right) \leq \frac{\mathbb{E}(X)}{\epsilon}$$

This result is known as Markov's inequality.

Proof. Note that $\Pr\left(X \geq \epsilon\right) = \mathbb{E}(\mathbf{1}\{X \geq \epsilon\})$, where

$$\mathbf{1}\{X \geq \epsilon\} = \begin{cases} 1 & \text{if } X \geq \epsilon \\ 0 & \text{if } X < \epsilon \end{cases}$$

We have

$$\mathbf{1}\{X \geq \epsilon\} \leq \frac{X}{\epsilon}$$

for any $\epsilon > 0$, since X is non-negative. Finally, taking the expectation of both sides yields the desired result. ■

Proof of Weak LLN

Chebyshev's inequality

Let X be a random variable, and $\epsilon > 0$ be a positive constant. Then,

$$\Pr\left(|X-\mu| \geq \epsilon\right) \leq \frac{\mathbb{V}(X)}{\epsilon^2}$$

This result is known as Chebyshev's inequality.

Proof. Noting the equivalence of the two events

$$|X-\mu| \geq \epsilon \iff |X-\mu|^2 \geq \epsilon^2,$$

it holds that

$$\Pr\left(|X-\mu| \geq \epsilon\right) = \Pr\left(|X-\mu|^2 \geq \epsilon^2\right).$$

Then, by Markov's inequality, we have

$$\Pr\left(|X-\mu|^2 \geq \epsilon^2\right) \leq \frac{\mathbb{E}(|X-\mu|^2)}{\epsilon^2} = \frac{\mathbb{V}(X)}{\epsilon^2}. \quad \blacksquare$$

Proof of Weak LLN

Now we are ready to prove Weak LLN by showing

$$\lim_{n\to\infty} \Pr\left(|\overline{X}_n - \mu| \geq \epsilon\right) = 0 \text{ for any } \epsilon > 0.$$

$\boxed{\text{Proof of Weak LLN}}$

Step 1. By Chebyshev's inequality and $\mathbb{E}(\overline{X}_n) = \mu$, for any $\epsilon > 0$,

$$\Pr\left(|\overline{X}_n - \mu| \geq \epsilon\right) \leq \frac{\mathbb{V}(\overline{X}_n)}{\epsilon^2}$$

Step 2. Since $\{X_1, \dots, X_n\}$ are independently drawn from the same distribution,

$$\mathbb{V}(\overline{X}_n) = \frac{\mathbb{V}(X)}{n}.$$

Proof of Weak LLN

Step 3. Combining these results, we have

$$\Pr\left(|\overline{X}_n - \mu| \geq \epsilon\right) \leq \frac{\mathbb{V}(X)}{n\epsilon^2}$$

Taking the limit of both sides as $n \to \infty$,

$$\lim_{n\to\infty} \Pr\left(|\overline{X}_n - \mu| \geq \epsilon\right) \leq 0.$$

Since any probability cannot be less than zero, the above inequality holds with equality. ■

What if the data are not independent?

If the data are dependent, $\mathbb{V}(\overline{X}_n) = \mathbb{V}(X)/n$ does not hold. Thus, $\mathbb{V}(\overline{X}_n)$ does not necessarily converge to zero.

Supplementary Discussion

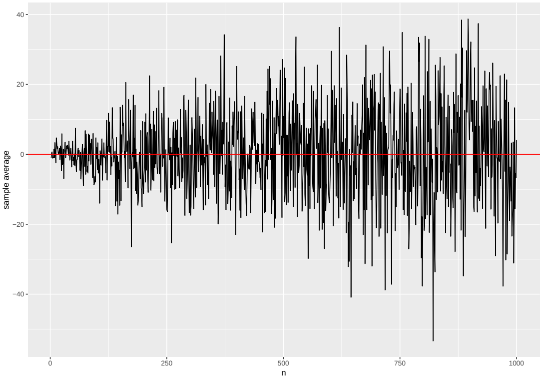
Dependent Data

- DGP 1: $X_1 \sim N(0,1)$, $X_i = X_{i-1} + N(0,1)$ for $i = 2, 3, \ldots$
- $\mathbb{E}[X_i] = 0$ for all i, but $\lim_{n\to\infty} \mathbb{V}(\overline{X}_n) \neq 0$.

```
>   sim <- function(n){
+     X <- rnorm(1)
+     for(i in 2:n) X <- c(X, X[i-1] + rnorm(1))
+     mean(X)
+   }
>   N <- 1000
>   R <- numeric(N-1)
>   for(i in 2:N) R[i-1] <- sim(i)
```

Dependent Data

```
> data <- data.frame(n = 2:N, mean = R)
> ggplot(data, aes(x = n, y = mean)) +
+   geom_line() + ylab("sample average") +
+   geom_abline(intercept = 0, slope = 0, color = "red")
```



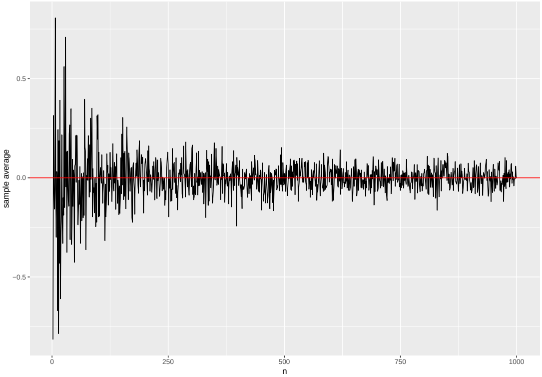
Dependent Data

- DGP 2: $X_1 \sim N(0,1)$, $X_i = 0.3X_{i-1} + N(0,1)$ for $i = 2, 3, \ldots$
- $\mathbb{E}[X_i] = 0$ for all i, and $\underline{\mathbb{V}(\overline{X}_n) \to 0}$.

```
>   sim <- function(n){
+     X <- rnorm(1)
+     for(i in 2:n) X <- c(X, 0.3*X[i-1] + rnorm(1))
+     mean(X)
+   }
>   N <- 1000
>   R <- numeric(N-1)
>   for(i in 2:N) R[i-1] <- sim(i)
```

Dependent Data

```
>   data <- data.frame(n = 2:N, mean = R)
>   ggplot(data, aes(x = n, y = mean)) +
+     geom_line() + ylab("sample average") +
+     geom_abline(intercept = 0, slope = 0, color = "red")
```



Estimation of conditional expectation

- Consider a randomized experiment setup: for example, $Y =$ blood pressure, $X = 1$ (new drug), $X = 0$ (placebo).
- The average treatment effect of the drug is measured by

$$\mathbb{E}[Y|X=1] - \mathbb{E}[Y|X=0]$$

- We can estimate $\mathbb{E}[Y|X=1]$ and $\mathbb{E}[Y|X=0]$ simply by computing the corresponding group-wise average: namely,

$$\frac{1}{\sum_{i=1}^{n} \mathbf{1}\{X_i = 1\}} \sum_{i=1}^{n} Y_i \mathbf{1}\{X_i = 1\} \xrightarrow{p} \mathbb{E}[Y|X=1]$$
$$\frac{1}{\sum_{i=1}^{n} \mathbf{1}\{X_i = 0\}} \sum_{i=1}^{n} Y_i \mathbf{1}\{X_i = 0\} \xrightarrow{p} \mathbb{E}[Y|X=0]$$

under similar conditions as above.

Estimation of conditional expectation

Treatment group: $X = 1$

Treatment group's average

$$\bar{Y}_{n1} \xrightarrow{p} \mathbf{E}[Y|X = 1]$$

Control group: $X = 0$

Control group's average

$$\bar{Y}_{n0} \xrightarrow{p} \mathbf{E}[Y|X = 0]$$