L.4: Central Limit Theorem

Econometrics 1: ver. 2024 Fall Semester

Naoki Awaya

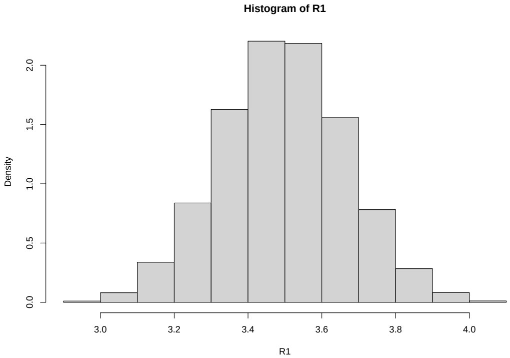
Central Limit Theorem: Introduction

Simulations

- In this simulation, we repeatedly compute the average of 100 dice rolls 10000 times, and plot the results in a histogram.
- Run the following **R** code:

```
>   library(extraDistr)
>   library(tidyverse)
>   Dice <- function(n) rdunif(n, 1, 6) %>% mean()
>   R1 <- numeric(10000)
>   for(i in 1:10000) R1[i] <- Dice(100)
>   hist(R1, freq = FALSE)
```

Simulations



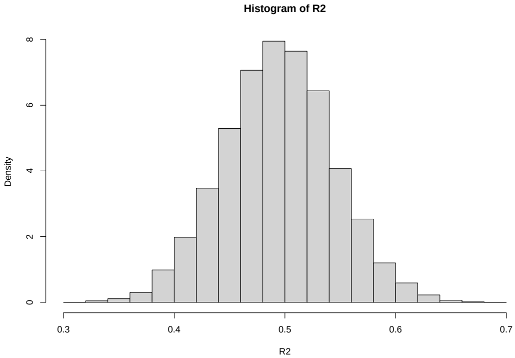
- The histogram looks almost **symmetric** and **unimodal** with its peak at $\mathbb{E}(X) = 3.5$.

Simulations

- Next, we perform a coin flipping experiment: head $= 1$, tail $= 0$.
- Similarly as above, compute the average of 100 coin flipping results 10000 times, and plot them in a histogram.

```
>   Coin <- function(n) rdunif(n, 0, 1) %>% mean()
>   R2 <- numeric(10000)
>   for(i in 1:10000) R2[i] <- Coin(100)
>   hist(R2, freq = FALSE)
```

Simulations



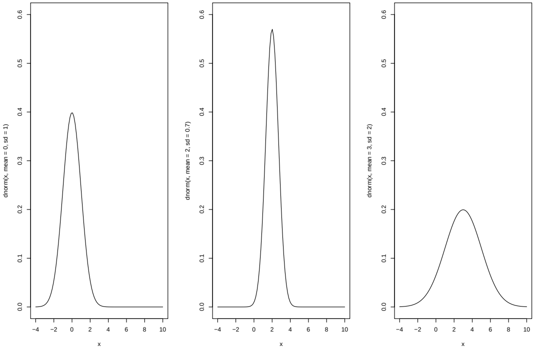
- Again, the histogram looks almost **symmetric** and **unimodal** with its peak at $\mathbb{E}(X) = 0.5$.

Normal distribution

- The most important probability distribution in the entire field of statistics is the normal distribution.
- The normal distribution is **symmetric** and **unimodal**.
- The shape of the normal distribution is fully characterized by two parameters: the mean μ and the standard deviation σ.
- The mean μ determines the **center** of the distribution, and the standard deviation σ determines the **width** of the curve.

Normal distribution

```
>   par(mfrow = c(1,3)) # split the graphic window into (1,3)
>   curve(dnorm(x, mean = 0, sd = 1),   xlim = c(-4,10), ylim = c(0,0.6))
>   curve(dnorm(x, mean = 2, sd = 0.7), xlim = c(-4,10), ylim = c(0,0.6))
>   curve(dnorm(x, mean = 3, sd = 2),   xlim = c(-4,10), ylim = c(0,0.6))
```



Normal distribution

- The normal distribution with mean μ and standard deviation σ is denoted as $N(\mu, \sigma^2)$, and its PDF is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- If X follows the normal distribution $N(\mu, \sigma^2)$, the probability that X is less than or equal to a is

$$\Phi(a; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{a} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx,$$

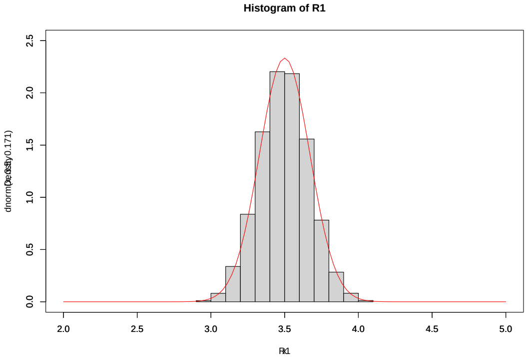
  where $\Phi(\cdot; \mu, \sigma)$ is the CDF of $N(\mu, \sigma^2)$.

- In particular, the normal distribution $N(0, 1)$ is referred to as the standard normal distribution.

Central Limit Theorem: Introduction

- It seems possible to approximate the histogram of the average dice rolls by an "appropriately chosen" normal distribution.

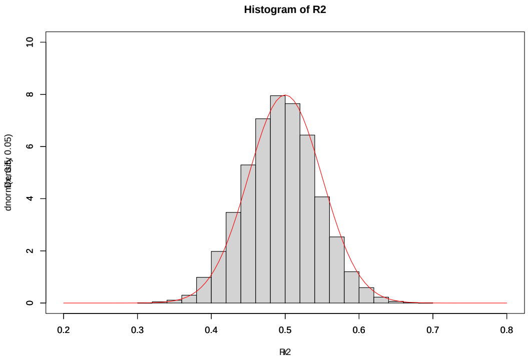
```
>   hist(R1, freq = FALSE, xlim = c(2,5), ylim = c(0,2.5))
>   par(new = T)
>   curve(dnorm(x, 3.5, 0.171), xlim = c(2,5), ylim = c(0,2.5), col = "red"
```



Central Limit Theorem: Introduction

- The same is true for the coin flipping experiment.

```
>   hist(R2, freq = FALSE, xlim = c(0.2,0.8), ylim = c(0,10))
>   par(new = T)
>   curve(dnorm(x, 0.5, 0.05), xlim = c(0.2,0.8), ylim = c(0,10), col = "re
```



Central Limit Theorem: Introduction

- Dice rolling and coin flipping experiments have different probability distributions, but the distribution of the sample mean can be approximated by a normal distribution in both cases.
- This surprising result is due to the central limit theorem.

CLT: Informal statement

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

$$\text{Pop mean of } X\text{: } \mu = \mathbb{E}(X), \quad \text{Pop variance of } X\text{: } \sigma^2 = \mathbb{V}(X),$$

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

Then, if n is sufficiently large, the probability distribution of $\overline{X}_n$ can be approximated by the normal distribution $N(\mu, \sigma^2/n)$.

Convergence in distribution

Convergence in distribution

Convergence in distribution

A sequence of random variables $(v_n)_{n=1}^{\infty}$ is said to converge in distribution to $N(\mu, \sigma^2)$ if

$$\lim_{n\to\infty} \underbrace{\Pr(v_n \leq x)}_{\text{CDF of } v_n} = \Phi(x; \mu, \sigma) \quad \text{for any } x.$$

- When $(v_n)_{n=1}^{\infty}$ converges to $N(\mu, \sigma^2)$ in distribution, we write

$$v_n \xrightarrow{d} N(\mu, \sigma^2)$$

- The followings are equivalent:

$$v_n \xrightarrow{d} N(\mu, \sigma^2), \quad v_n - \mu \xrightarrow{d} N(0, \sigma^2), \quad \frac{(v_n - \mu)}{\sigma} \xrightarrow{d} N(0, 1)$$

- μ and σ are the mean and standard deviation of v_n, respectively.

Convergence in distribution

- A formal statement of CLT is the following: under the aforementioned assumptions,

$$\underbrace{\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}}_{\text{standardized sample mean}} \xrightarrow{d} N(0,1)$$

(* Note that $\mathbb{E}(\overline{X}_n) = \mu$ and $\mathbb{V}(\overline{X}_n) = \sigma^2/n$.)

- What is remarkable for this result is that the distribution of X can be (almost) anything.
- The proof of CLT is VERY complicated, and thus is omitted.

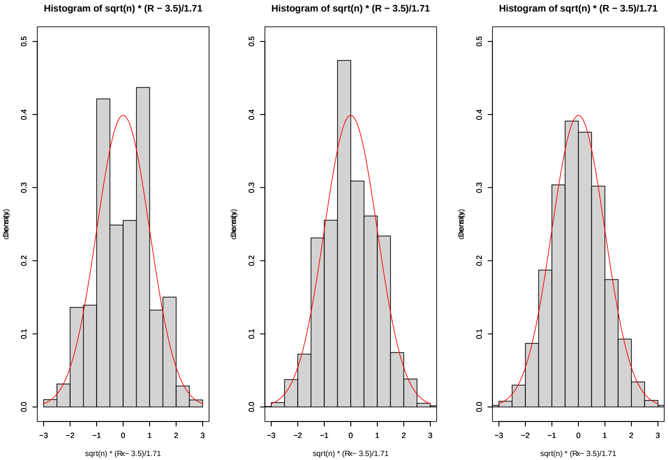
Convergence in distribution

Dice roll simulation:

- $\mathbb{E}[X] = 3.5$
- $\mathbb{V}[X] = \sum_{i=1}^{6} (i - 3.5)^2/6 = 35/12$, $\sqrt{\mathbb{V}[X]} \approx 1.71$.

```
>   plot_Dice <- function(n){
+     R <- numeric(10000)
+     for(i in 1:10000) R[i] <- Dice(n)
+     hist(sqrt(n)*(R - 3.5)/1.71, freq = FALSE, xlim = c(-3,3), ylim = c(0, 0.5))
+     par(new = T)
+     curve(dnorm(x), xlim = c(-3,3), ylim = c(0, 0.5), col = "red")
+   }
>   par(mfrow = c(1,3))
>   plot_Dice(3);  plot_Dice(8);  plot_Dice(500)
```

Convergence in distribution



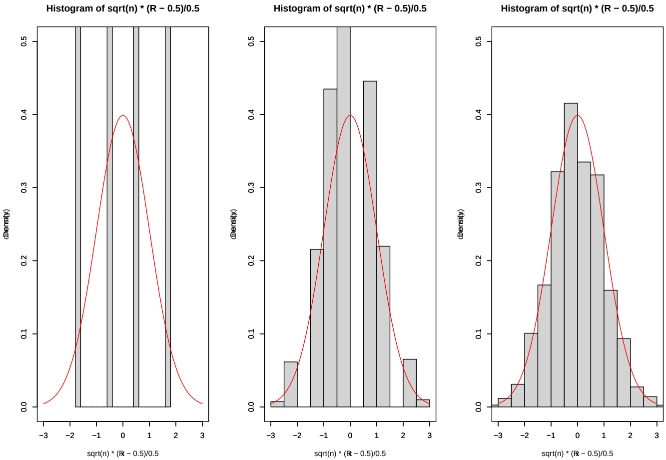
Convergence in distribution

Coin toss simulation:

- $\mathbb{E}[X] = 0.5$
- $\mathbb{V}[X] = (0 - 0.5)^2/2 + (1 - 0.5)^2/2 = 0.25,\ \sqrt{\mathbb{V}[X]} = 0.5.$

```
>   plot_Coin <- function(n){
+     R <- numeric(10000)
+     for(i in 1:10000) R[i] <- Coin(n)
+     hist(sqrt(n)*(R - 0.5)/0.5, freq = FALSE, xlim = c(-3,3), ylim = c(0, 0.5))
+     par(new = T)
+     curve(dnorm(x), xlim = c(-3,3), ylim = c(0, 0.5), col = "red")
+   }
>   par(mfrow = c(1,3))
>   plot_Coin(3);  plot_Coin(8);  plot_Coin(500)
```

Convergence in distribution



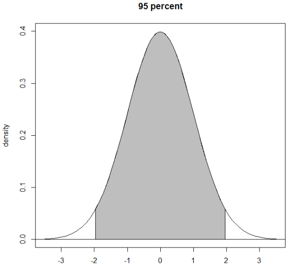
Confidence Interval

Confidence Interval

Let Z be distributed as the standard normal $N(0,1)$. Then, it holds that

$$\Pr(-1.96 \leq Z \leq 1.96) = 0.95 \text{ (approximately)}$$

That is, the area of grayed part in the following figure is equal to 0.95



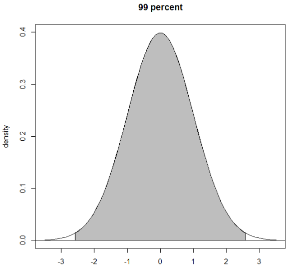
An intuition: when drawing a random number from $N(0,1)$, the value will be included in $[-1.96, 1.96]$ about 95 times out of 100.

Confidence Interval

Similarly, the 99% interval for the standard normal Z is given by

$$\Pr(-2.58 \leq Z \leq 2.58) = 0.99 \text{ (approximately)}$$

That is, the area of grayed part in the following figure is equal to 0.99



Convergence in distribution

```r
>   integrate(dnorm, -1.96, 1.96)
```

```
## 0.9500042 with absolute error < 1e-11
```

```r
>   integrate(dnorm, -2.58, 2.58)
```

```
## 0.99012 with absolute error < 1.9e-08
```

Confidence Interval

- Let $\{X_1, \dots, X_n\}$ be a random sample of n observations. According to the CLT, the distribution of the standardized sample mean

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}$$

can be approximated by $N(0, 1)$ as n increases.

- Thus, we have

$$\Pr\left(-1.96 \leq \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \leq 1.96\right) \approx 0.95.$$

for sufficiently large n.

Confidence Interval

- Hence, if n is sufficiently large,

$$
\begin{aligned}
0.95 &\approx \Pr\left(-1.96\frac{\sigma}{\sqrt{n}} \le \overline{X}_n - \mu \le 1.96\frac{\sigma}{\sqrt{n}}\right)\\
&= \Pr\left(-\overline{X}_n - 1.96\frac{\sigma}{\sqrt{n}} \le -\mu \le -\overline{X}_n + 1.96\frac{\sigma}{\sqrt{n}}\right)\\
&= \Pr\left(\overline{X}_n - 1.96\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X}_n + 1.96\frac{\sigma}{\sqrt{n}}\right)
\end{aligned}
$$

- This implies that the population mean μ is included in the interval

$$
\left[\overline{X}_n - 1.96\frac{\sigma}{\sqrt{n}},\ \ \overline{X}_n + 1.96\frac{\sigma}{\sqrt{n}}\right]
$$

with approximately 95% probability. This interval is called the 95% confidence interval (CI) of μ.

Confidence Interval

Even though we do not know the true value of μ, the interval in which μ is contained with a certain probability is computable.

This is the power of the central limit theorem!

- The length of the CI indicates the precision of inference, which is inversely proportional to $\sqrt{n}$.
  - E.g., the length of the 95% CI is $3.92\sigma/\sqrt{n}$.
  - If you want to halve the CI, you need to increase the sample size four-fold (not two).
- Similarly as above, the 99% CI can be computed by

$$\left[\overline{X}_n - 2.58\sigma/\sqrt{n},\ \ \overline{X}_n + 2.58\sigma/\sqrt{n}\right]$$

Simulation

- In the dice roll experiment, the 95% CI is $\left[\overline{X}_n - \frac{3.35}{\sqrt{n}},\ \ \overline{X}_n + \frac{3.35}{\sqrt{n}}\right]$.

```
>   CI <- function(n) Dice(n) + c(-3.35, 3.35)/sqrt(n)
```

- We compute the CI 10000 times and how often μ $(= 3.5)$ fall inside the interval for different n's.

```
>   test <- function(n){
+     result <- numeric(10000)
+     for(i in 1:10000){
+       CI <- CI(n)
+       result[i] <- ifelse(CI[1] < 3.5 & 3.5 < CI[2], 1, 0)
+     }
+     mean(result)
+   }
```

Simulation

```
> test(5)
```

```
## [1] 0.9328
```

```
> test(10)
```

```
## [1] 0.9433
```

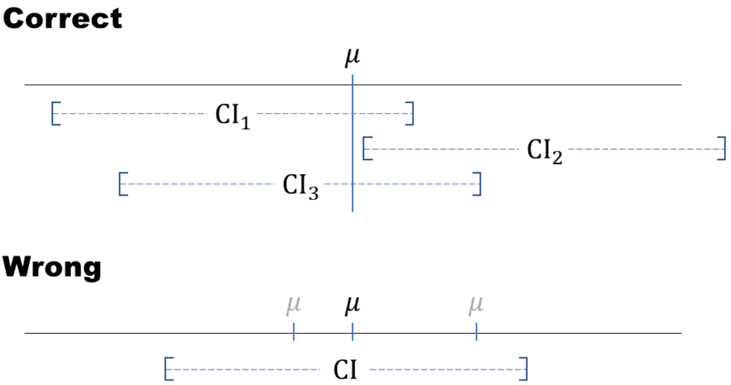
```
> test(1000)
```

```
## [1] 0.9505
```

- Thus, as n gets larger, the probability of $\mu \in 95\%$ CI becomes closer to the nominal level (95%).

A common misunderstanding

- μ is a fixed parameter (not a r.v.), and what is random is the CI.



t-distribution

t-distribution

- In order to compute the CI, we need to know the population standard deviation σ, which is typically unknown.
- Thus, in practice, we replace σ by the sample standard deviation $\hat{\sigma}_n$:

$$\hat{\sigma}_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}$$

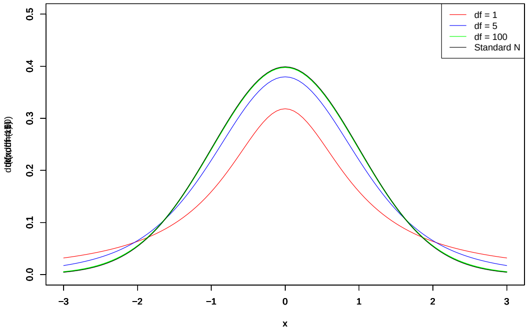
and estimate the CI by

$$\widehat{CI}_n = \left[\overline{X}_n - 1.96\frac{\hat{\sigma}_n}{\sqrt{n}},\ \ \overline{X}_n + 1.96\frac{\hat{\sigma}_n}{\sqrt{n}}\right]$$

- When σ is replaced by $\hat{\sigma}_n$, $\sqrt{n}(\overline{X}_n - \mu)/\hat{\sigma}_n$ does not distribute as $N(0,1)$, but is distributed as a t-distribution with $n-1$ degrees of freedom.

t-distribution

- The t-distribution has a similar shape as the standard normal distribution, and converges to $N(0,1)$ as n increases.



- Thus, as long as n is sufficiently large, we can treat $\sqrt{n}(\overline{X}_n - \mu)/\hat{\sigma}_n$ as the standard normal r.v. and $\widehat{CI}_n$ serves as a valid CI.