R 語言教學

鍾旻錡, 陳柏瑜

Statistics with Recitation

NTU Econ

2020.11.11

Outline

1. Sampling, Asymptotic Distributions & CLT
2. Introduction to tidyverse
3. Data Manipulation

Sampling, Asymptotic Distributions & CLT

WLLN: Convergence of Sample Mean

Given that $\{X_i\}_{i=1}^{n} \sim^{i.i.d} (E(X_1), Var(X_1))$

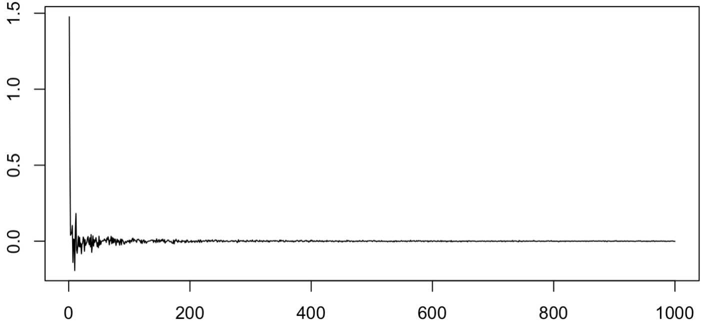
When $r = 1$, the convergence of our sample moment is the following:

Suppose the population distribution is $N(0, 1)$

$$\frac{1}{n}\sum_{i=1}^{n} {X_i}^r \xrightarrow{p} E({X_1}^r)$$

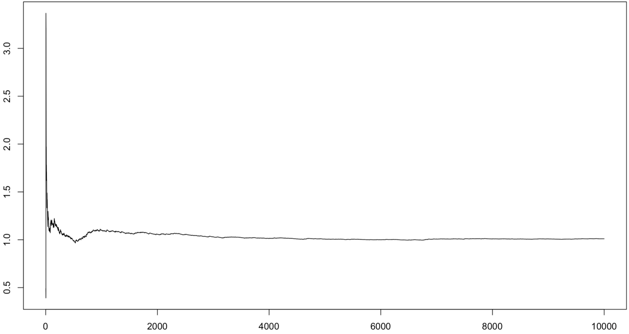
WLLN: Convergence of Sample Mean

$$\frac{1}{n}\sum_{i=1}^{n} X_i{}^r \xrightarrow{p} E(X_1{}^r)$$



WLLN: Convergence of Sample Variance

Similarly, we can look at the convergence of variance (Suppose the population distribution is $N(0,1)$)



WLLN: Practice

Practice in R

Given that the population distribution is $Binomial(n = 10, p = 0.2)$

Find the plot of convergence

Sampling Distributions

Practice in R:

Given $\{X_i\}_{i=1}^{n} \sim^{i.i.d} N(\mu, \sigma^2)$

Use X and follow the process below:

1. generate random vectors
2. graph the histogram
3. add density plot onto the histogram

Apply the above process on the random variable listed in next slides.

Sampling Distributions

- $Z_1 = \frac{X_1 - \mu}{\sigma} \sim N(0, 1)$
- $\chi_1 = Z_1^2 \sim \chi^2(1)$
- $\chi_1 + \chi_2 = Z_1^2 + Z_2^2 \sim \chi^2(2)$
- $t = \frac{Z_1}{\sqrt{\chi_2/1}} \sim t(1)$
- $F = \frac{\chi_1/1}{(\chi_2+\chi_3)/2} \sim F(1, 2)$

CLT

Recall that $\bar{X}_n$ is the main character.

CLT states that

$$\frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim^d N(0,1)$$

as $n \to \infty$

Let's generate a bunch of $\bar{X}_n$ to look at the distribution of standardized $\bar{X}_n$!

Introduction to tidyverse

Pipeline

tidyverse contains a bunch of packages such as:

- ggplot2
- dplyr
- stringr
- and pipeline %>%
- Visit the website: https://www.tidyverse.org

Pipeline as an operator

In R, we can easily define an operator by assigning a function to **‘SomeOperatorName‘**. You can also rewrite the operator in base function, such as **‘+‘**

```
`+` <- function(e1, e2) {
  if (is.character(e1) | is.character(e2)) {
    paste0(e1, e2)
  } else {
    base::`+`(e1, e2)
  }
```

Pipeline - Graphing Example

Before adding pipeline

```
x <- rnorm(10000, 0, 1)
y <- pnorm(x, 0, 1)
hist(y)
```

After adding pipeline

```
rnorm(10000, 0, 1) %>%
pnorm(0, 1) %>%
hist()
```

Data Manipulation

Efficient Frontier

You have an asset that gives you some return randomly.

Let R_1 be a random variable, denoting the asset return.

What will you expect about R_1 ?

You may care about:

- $E(R_1)$, $Var(R_1)$
- It seems like $E(R_1)$ tends to be higher when $Var(R_1)$ is relatively high.

Is it true?

Efficient Frontier - Risk Diversification

Efficient Frontier - Risk Diversification

If you have more than one assets that are not identical, then the returns of different assets vary.

In this case, there might be some chance to lower down your risk, i.e. you only need to bear a lower risk while earning the same amount of expected return.

Efficient Frontier - Portfolio

Now consider two risky assets that gives you return R_1 and R_2, which are two different random variables.

Say you want to make a portfolio consisting of the two assets, how will you choose the weight ω?

$$P = \omega R_1 + (1 - \omega) R_2$$

Efficient Frontier - Portfolio

Think

- some ω may give me a portfolio with least variance(risk) in return
- some ω may provide a portfolio that is worse than any of the two assets

Efficient Frontier - Portfolio Return and Risk

Given $R_1 \sim (\mu_1, \sigma_1^2)$, $R_2 \sim (\mu_2, \sigma_2^2)$

$$P = \omega R_1 + (1 - \omega) R_2$$

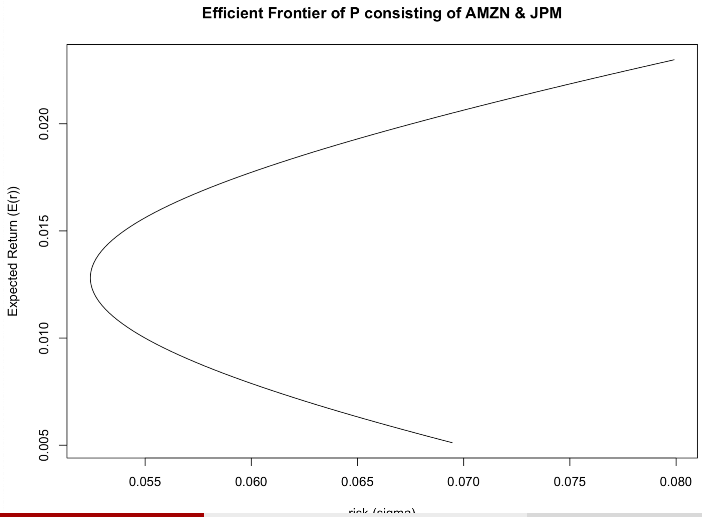
- $E(P) = \omega E(R_1) + (1 - \omega) E(R_2)$
- $Var(P) = \omega^2 Var(R_1) + (1 - \omega)^2 Var(R_2) + 2Cov(R_1, R_2)$

Efficient Frontier - Practice with Yahoo Finance

Let's go to Yahoo Finance and download real stock price data.

- https://finance.yahoo.com

Efficient Frontier - Portfolio Return and Risk



感謝大家聆聽