Financial Econometrics

Lecture I

Probabilistic and Statistical Review

Narapong Srivisal, Ph.D.

What is Econometrics?

EconoMetrics

Economics + Metrics

(a standard of measurement)

apply statistics

Method to analyse economic data

What is Econometrics?

- ***Merriam-Webster***: the application of statistical methods to the study of economic data and problems.
- ***Oxford***: The branch of economics concerned with the use of mathematical methods (especially statistics) in describing economic systems.
- ***Cambridge***: the part of economics that uses mathematics and statistics to understand financial and economic information

finance is a subset of economic

What to Expect from this Course?

α & β change all the time
How a stock exposed to market risk

- A lot of **statistics** and **mathematics**.
- Applied to economic or financial data to estimate some kinds of relationships: ~~correlation~~ → this is more complicated | CAPM → excess return $= \alpha + \beta$ [market excess return]

  abnormal return if $\neq 0$

  → Fa a-French ; add size, BV for explanatory

  → Carhart

  How can we forecast return
  - **What factors effect returns?**
  - **What are the "betas" of a given stock?**

    ค.ผันผวนของตลาด risk ; model to estimate
  - **What are the volatilities of a given stock?**

    ↳ can change any time
- The goals are to understand:
  - **Different data structures.**
    - cross-section ; ignore time, many individuals
    - time-series data ; 1 individual in different time
    - panel data ; many individuals in different times

    tool to estimate ; suitable estimator for different data structure
  - **When and which estimators should be used.**
  - **How to interpret the estimates.**

    ↳ translate result in simple words
  - **How to make statistical inference.**

    ↳ how to translate sample to population
    - confidence interval
    - hypothesis test

Outline

Probabilistic Review

- **Random Variable**
  - Distribution
  - Expectation
  - Variance
  - Other Moments
- **Random Vector**
  - Distribution
  - Expectation Revisit
  - Law of Iterated Expectation
  - Independence
  - Mean Independence
  - Covariance and Correlation

MSF Chula

CHULALONGKORN BUSINESS SCHOOL
FLAGSHIP FOR LIFE

Random Variable

we don't know outcome in advance → uncertain

- Probability theory begins with the idea of "experiment," meaning any process whose outcome is not known for certain in advance.
- **Sample space** is the set of all possible outcomes of an experiment. can be measured is number
- E.g. The experiment: tossing a fair coin twice.
  Sample space: $\{TT, TH, HT, HH\}$ → cannot be measured → translate to random variable
- A **random variable** usually means a function from the sample space to the set of real number. It takes on numerical values which are determined by an experiment.
- E.g. A random variable X for the tossing a fair coin experiment may be the number of heads showing up: $X \in \{0, 1, 2\}$
- E.g. A random variable W for the tossing a fair coin experiment may be whether there's a head showing up at all: $W \in \{0, 1\}$

describe random variable

Distribution

⇒ ① What are value that random variable can take
② Chance of happening

- Since the outcome of an experiment is random, the value of X is also random. The **distribution of a random variable X** is a description of the **probability** that X takes on different values.

- E.g. A random variable for the tossing a fair coin experiment may be the number of heads showing up.

$\Pr\{X \geq 1\}$ → add up all prob in set

4 cases:

$$\{TT \Rightarrow X = 0, TH \Rightarrow X = 1, HT \Rightarrow X = 1, HH \Rightarrow X = 2\}$$

①

So, X can take values $0, 1, 2$. Then, the distribution of X is the probabilities that $X = 0$, $X = 1$, and $X = 2$, which respectively are

PMF of X : $X = \begin{cases} 0 & p = 1/4 \\ 1 & p = 1/2 \\ 2 & p = 1/4 \end{cases}$

②

$$p(0) = \frac{1}{4}, p(1) = \frac{1}{2}, \text{ and } p(2) = \frac{1}{4}$$

- E.g. The distribution of W in earlier example is $p(0) = \frac{1}{4}, p(1) = \frac{3}{4}$

$E[X] = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1$

$E[X^2] = 0^2 \cdot \frac{1}{4} + 1^2 \cdot \frac{1}{2} + 2^2 \cdot \frac{1}{4} = 1\frac{1}{2}$

$E[2X] = 2E[X] = 2 \cdot 1 = 2$ linear transformation

$E[2+X] = 2 + E[X] = 3$

Discrete Distribution

each point has prob, mass ⇒ PMF

- A discrete random variable takes on only finitely many (or at most countably many) different values.
- Let $x_1, \dots, x_k$ donote the finitely many values that X can take on, then the distribution is completely described by the probability with which X takes each value, or the **probability mass function (PMF):**

$$p(x) = P\{X = x\}$$

- Properties of the p.m.f:

  can be distribution

  (i) $p(x) \geq 0$ for all x

  (ii) $\Sigma_{i=1}^{k} p(x_i) = 1$ sum of all cases

  (iii) $P\{X \in A\} = \Sigma_{i=1}^{k} I\{x_i \in A\} p(x_i)$

- E.g. X is a Bernoulli random variable with parameter p if it equals 1 with probability p and equal 0 with probability $1 - p$. We use the notation $X \sim$ **Bernoulli(p)**

$X \sim N(0,1)$

notation / name of random variable

has distribution as

$X \sim$ Bernoulli (0.7)

prob that x equals to 1

$$X \begin{cases} 1 \Rightarrow P(X=1) = 0.7 \\ 0 \Rightarrow P(X=0) = 0.3 \end{cases}$$

Continuous Distribution

$P(X=0.1) = \frac{1}{\infty}$

$= 0$

Prob of particular value is always $= 0$

- A continuous random variable takes on a continuum of values, such as all the Real numbers or the interval [0,1].

each point doesn't have mass

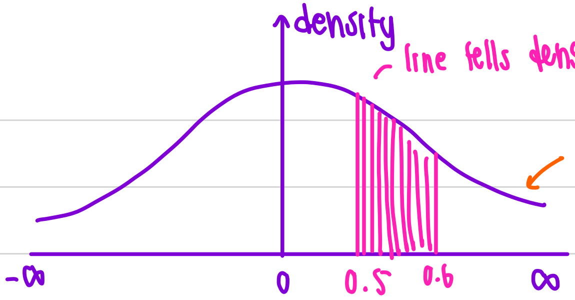
PDF

- Since there are infinitely many values that a continuous random variable can take, the probability that a specific value can take is theoretically zero.
- The distribution of a continuous random variable is completely described by the **probability density function (PDF)**, denoted by $f(x)$, which satisfies the following properties:

(i) $f(x) \geq 0$ for all x

(ii) $\int_{-\infty}^{\infty} f(x)dx = 1$

(iii) $P\{X \in A\} = \int_{-\infty}^{\infty} I\{x \in A\} f(x)dx$



line tells density of each point

probability density function (PDF) → we're talking about prob of range

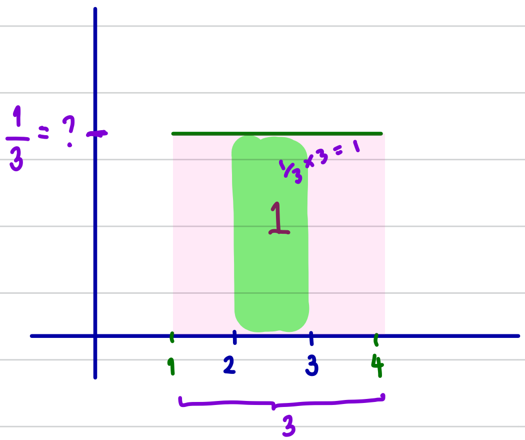
value of random variab

$Pr\{x \in [0.5, 0.6]\}$ → area from 0.5 to 0.6 under pdf curve → have to add up all density within range

$= \int_{0.5}^{0.6} pdf(x)\,dx$

↳ sum up all point within range

$X \sim$ UNIFORM $[1,4]$



$P(x < 2) = \frac{1}{3}$

$P(x \leq 2) = \frac{1}{3}$; because $P(x = 2) = 0$ for continuous r.v.

$P(x \leq 2) = P(x<2) + P(x=2)^{\to 0}$

$= P(x<2)$

$P(x < 3) = \frac{2}{3}$

$P(x > 3) = \frac{1}{3}$

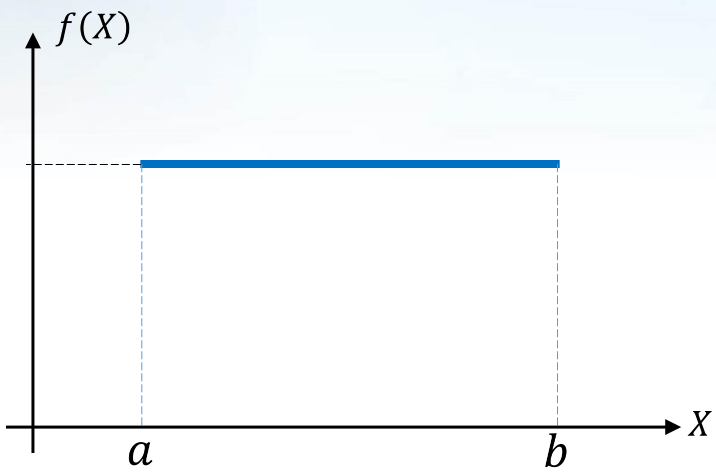
$P(x \in (2,3)) = \frac{1}{3}$

↳ $= P(x<3) - P(x<2)$ (cumulative prob)

$= \frac{2}{3} - \frac{1}{3} = \frac{1}{3}$

- E.g. X is uniformly distributed on $[a, b]$, or $X{\sim}U(a, b)$, the p.d.f. is

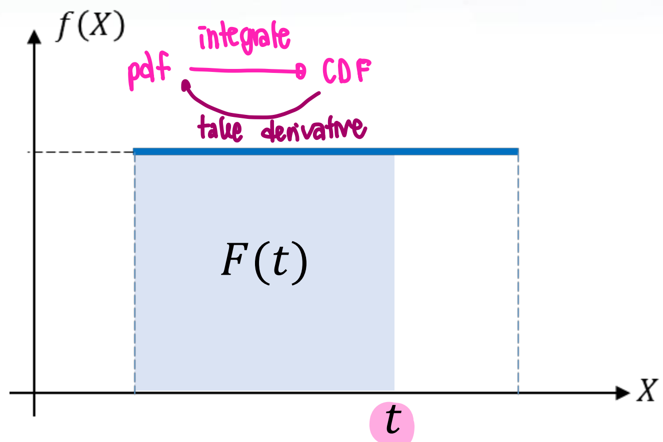
$$f(X) = \begin{cases} \frac{1}{b-a} & if\ a \le X \le b \\ 0 & otherwise \end{cases}$$

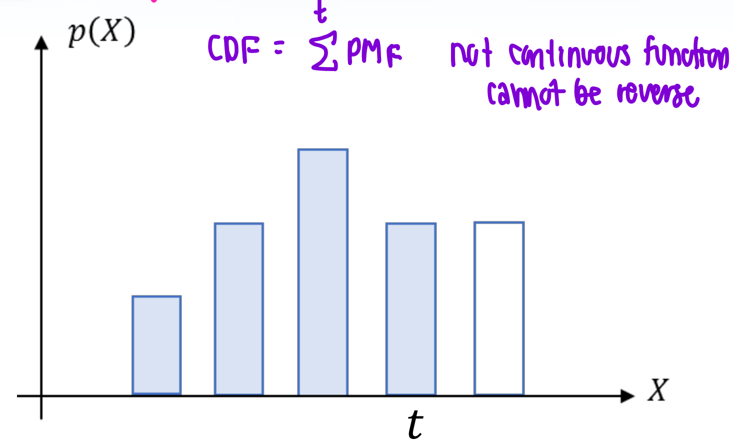


- A function of a random variable is also a random variable.
- Some random variables are not either discrete or continuous. E.g. $g(X) = \max\{X, 0\}$ and $X{\sim}U(-1,1)$ or a distribution of value of an option. We call such random variable as a mixed discrete-continuous random variable.

Cumulative Distribution Function

- The **cumulative distribution function (CDF)** of a random variable X is simply the function $F(t) = P\{X \leq t\}$ $= \int_{-\infty}^{t} \text{pdf}(x)\,dx$





- It is the integration of PDF or summation of PMF up to point t.
- Hence, for a continuous random variable, PDF is the derivative of CDF.
- $P\{a < X \leq b\} = P\{X \leq b\} - P\{X \leq a\} = F(b) - F(a)$

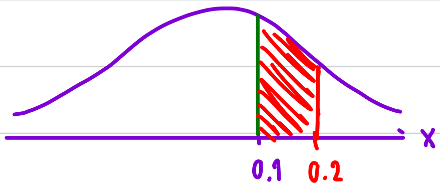
density of some particular value

$$\text{PDF} \xrightarrow{\text{integrate}} \text{CDF} \xleftarrow{\text{sum}} \text{PMF}$$

we restrict area
so there's no constant

we don't want to know the density of a particular value
but area of range
prob

$$\text{pdf} = \int_{0.1}^{0.2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx$$



z-table : CDF of standard normal

$$\int_a^b f(x)\, dx = F(x) + \cancel{C}$$

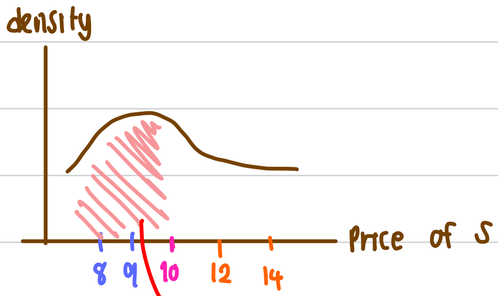
no constant

Example

derivative : financial product based on underlying asset
(has mixed distribution of discrete and continuous; e.g. warrent)

CALL OPTION : right but not obligation to buy an underlying asset (stock S) at some particular price (strike price)

↳ distribution of call option's price must link to the distribution of underlying asset's price



exercise price = $10

value of call option ‖ discrete at zero; continuous $(0, \infty)$

$\Pr\{\text{call option value} = 0\}$

$= \Pr\{\text{stock } S \text{ price} \leq 10\}$

0 → not exercise call option if stock price ≤ 10

4 (profit) ↳ chance of value of call option = 4 is equal to the chance that stock's price = 14

Expectation

middle value
≡ average value ≡ mean | center of distribution

- The **expected value** or **expectation** or **mean** of a random variable X is one measure of the "central tendency" of the distribution of X, typically denoted by $\boldsymbol{E[X]}$ or $\boldsymbol{\mu_X}$.
- If X is a discrete random variable with PMF $p(x)$ taking on values $x_1, \ldots, x_k$, then

all possible value of r.v.
prob of each happens

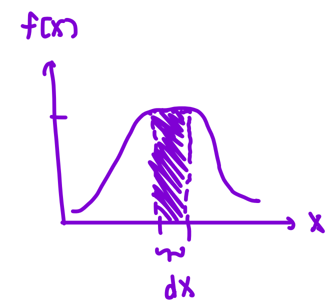
$$E[X] = \sum_{i=1}^{k} \boldsymbol{x_i p(x_i)}$$

- If X has a continuous distribution with p.d.f. $f(x)$, then

density of it (pdf)

$$E[X] = \int_{-\infty}^{\infty} \boldsymbol{x f(x)} d(x)$$

↳ sum all up



- Eg.: Let $X \sim U(a, b)$, then

$$E[X] = \int_a^b x \cdot \frac{1}{b-a} dx = \left(\frac{1}{b-a}\right) \int_a^b x dx = \left(\frac{1}{b-a}\right)\left(\frac{b^2 - a^2}{2}\right) = \frac{1}{2}(a+b)$$

Example : Grade : 4 4 4 3 4 3 2 2 2 2 4 1

average = $\overset{\#}{\frac{5}{12}}(4) + \frac{2}{12}(3) + \frac{4}{12}(2) + \frac{1}{12}(1)$.

prob
/chance
that student get 4

- Eg.: Let X be a random variable and let $A \subseteq \mathbb{R}$, then

$$E[I\{X \in A\}] = \int_{-\infty}^{\infty} I\{X \in A\} f(x)dx = P\{X \in A\}$$

- The expected value of a function of random variable, $g(X)$, is defined similarly:

plug in x^2 and multiply with prob

$$E[\overset{x^2}{g(X)}] = \sum_{i=1}^{k} \overset{x^2}{g(x_i)}p(x_i) \text{ , or } E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

- Eg.: Let X be a random variable and a, b be constant numbers, then

Σ for discrete

$$E[a + bX] = \int_{-\infty}^{\infty}(a + bx)f(x)dx$$

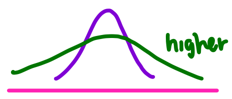
$$= \int_{-\infty}^{\infty} af(x)dx + \int_{-\infty}^{\infty} bxf(x)dx$$

$$= a\underbrace{\int_{-\infty}^{\infty} f(x)dx}_{=1} + b\underbrace{\int_{-\infty}^{\infty} xf(x)dx}_{=E[X]} = a + bE[X]$$

Variance

- The **variance** of a random variable X is one measure of the "dispersion" or "spread" of its distribution, denoted by $\boldsymbol{Var[X]}$ or $\boldsymbol{\sigma_X^2}$
- It is defined as the expected value of the squared deviation of X from its mean:

how far from mean $E[x^2]$ always $\geq (E[x])^2$


higher variance

$$Var[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

to determine both ⊖ and ⊕ from center of distribution

- $Var[X]$ is always greater than zero, unless X is just a constant. → variance = 0

hard to interpret

- The unit of $Var[X]$ is the square of the unit of X. Thus, the dispersion is sometimes measured by the square root of the variance, which is referred to as the **standard deviation**. → same unit with x
- Let X be a random variable and a, b be constant numbers, then

$$Var[a + bX] = b^2 Var[X]$$

Property of variance :

$Var(a + X) = E[[(a + X) - E(a + X)]^2]$

$\cancel{a} + x - \cancel{a} - Ex$

$= E[(x - E(x))^2] = Var[x]$

$Var[ax] = E[(ax - E(ax))^2]$

$= E[(a[x - E(x)])^2] = a^2 E[(x - E(x))^2] = a^2 Var(x)$

$sd(ax) = |a| \, sd(x)$

Moments

moment → expectation of anything in general
E[ ]

mean ; measure of center

- The k^{th} **(raw) moment** of a random variable is $E[X^k]$

2nd centered moment : Var[X]

- The k^{th} **centered moment** of a random variable is $E[(X - E[X])^k]$
- The k^{th} **standardized moment** of a random variable is

↳ from 3th moment onwards

$$\frac{E\left[(X - E[X])^k\right]}{\sigma_X^k}$$

standardized this to make it unitless

- The standardized moment is unit-less. Thus, it is helpful for comparing random variables with different units
- The third standardized moment is called **skewness**
- The fourth standardized moment is called **kurtosis**, measuring the likelihood of extreme values of random variables.

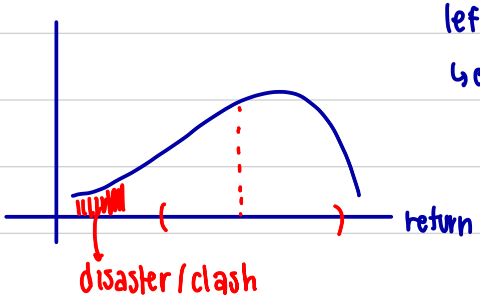
SKEWNESS : $\dfrac{E\left[(X - E(X))^3\right]}{\sigma_X^3}$

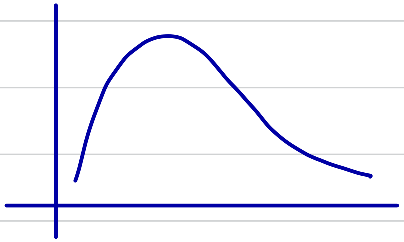
negative

left skewed negative

↳ e.g. there is possibility that return goes down a lot (very negative return)

↓ not much



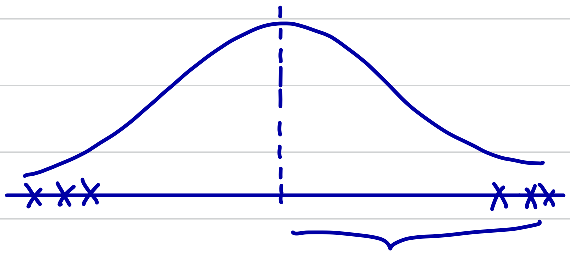


right skewed positive

KURTOSIS → 4^{th} std moment

$\dfrac{E\left[(X - \mu_X)^4\right]}{\sigma_X^4}$ always ⊕ ↑↑↑ very high

to capture extreme value



For normal distribution → kurtosis = 3

excess kurtosis = 0

more than 1 random variable

Random Vector

- The data we observe in the real world typically involve more than one variable.
- A **random vector** is simply a vector containing several random variables, i.e. a d-dimensional random vector a function from the sample space to $\mathbb{R}^d$
- The **joint distribution** of a random vector is just a description of the probability of the vector values that it takes.
- If we let $x_1, \dots, x_k$ be all the values that X can take, and $y_1, \dots, y_l$ be all the possible values for Y, then the joint distribution is completely described by the probability that each pair of value for (X, Y) can take, i.e. the PMF is

random variable by itself

$$p(x, y) = P\{X = x, Y = y\}$$

chance that each combination takes

(x,y) Distribution: values / chances it can take — discrete: prob; continuous: density

- Properties (i) $p(x, y) \geq 0$ for all (x, y)

  sum of all prob

  (ii) $\Sigma_{i=1}^{k} \Sigma_{j=1}^{l} p(x_i, y_j) = 1$

  (iii) $P\{(X, Y) \in A\} = \Sigma_{i=1}^{k} \Sigma_{j=1}^{l} I\{(x_i, y_j) \in A\} p(x_i, y_j)$

- Eg.: Suppose (X, Y) is a discrete random vector where X takes on values 1,2,3 and Y takes on values -2,0,2. The PMF of (X, Y) is summarized by

9 combinations

all of these generate distribution called joint distribution of (x, y)

prob of each combination joint probability

$X \backslash Y$	-2	0	2	$p(x)$
1	1/3	0	0	1/3 p(x=1)
2	0 +	1/3 +	1/6	1/2 p(x=2)
3	0 +	1/6 +	0	1/6 p(x=3)
$p(y)$	1/3	1/2	1/6	1

to tell that we have more than 2 r.v. but we want to focus on x

marginal distribution of x

↳ marginal probability

marginal distribution of y

Distribution / Probability for random vector → Expectation

marginal : $p(x)$, $p(y)$ → $E[X]$, $E[X^2]$, $E[g(y)]$ involves only 1 r.v.

joint : $p(xy)$ $\xrightarrow{\text{pg. 26}}$ $E[x+y]$, $E[xy]$, $E[x^2y]$, $E[g(x,y)]$ ↗ each combination of x, y that can happen then weighted with joint prob

conditional : $p(x|y)$ $\xrightarrow{\text{pg. 27}}$ $E[x|y]$

Conditional probability

$P(X=2 \mid y=0)$ strict attention on this group only

↳ conditional on or given $y=0$

What is the proportion / chance that $x=2$ ⇒ $= \dfrac{\frac{1}{3}}{\underbrace{0+\frac{1}{3}+\frac{1}{6}}_{\text{total}}} = \dfrac{p(x=2, y=0)}{p(y=0)}$ → $\dfrac{\text{joint}}{\text{marginal}}$

$X \backslash Y$	-2	0	2	$p(x)$
1	1/3	0	0	1/3
2	0	$\boxed{1/3}$	1/6	1/2
3	0	1/6	0	1/6
$p(y)$	1/3	1/2	1/6	1

think of $y=0$ is the total 100%.

return

$X\backslash Y$	-2 (project generate ⊖ return)	0 (0 return)	2 (⊕ return)	$p(x)$
1 (฿ 1bn project)	1/3	0	0	1/3
2	0	1/3	1/6	1/2
3	0	1/6	0	1/6
$p(y)$	1/3	1/2	1/6	1

size of project

$\frac{1}{3} = P(X=1)$ ⇒ the prob that we get project of size 1 billion

sample from whole population

$0 = \underbrace{P(X=1 \mid y=2)}$ ⇒ the prob that we get 1 billion project if the selected project is positive return

<we know that the selected projects give positive return>

$$\frac{P(X=1, y=2)}{P(y=2)} = \frac{0}{1/6} = 0$$

$P(X=2 \mid y=0) = 0$

$P(X \mid y=0)$

- $X=1 \Rightarrow P(X=1 \mid y=0) = ?$
- $X=2 \Rightarrow P(X=2 \mid y=0) = ?$
- $X=3 \Rightarrow P(X=3 \mid y=0) = ?$

this is function of X

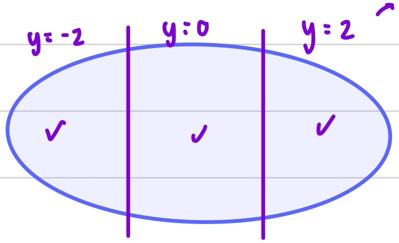
as we change X, the value of function change

(conditional) distribution of X given that $y=0$

$P(X=2 \mid y)$

- $y=-2 \Rightarrow P(X=2 \mid y=-2)$
- $y=0 \Rightarrow P(X=2 \mid y=0)$
- $y=2 \Rightarrow P(X=2 \mid y=2)$

function of y



we divide population

$X\backslash Y$	-2	0	2	$p(x)$
1	1/3	0	0	1/3
2	0	1/3	1/6	1/2
3	0	1/6	0	1/6
$p(y)$	1/3	1/2	1/6	1

list all possible cases

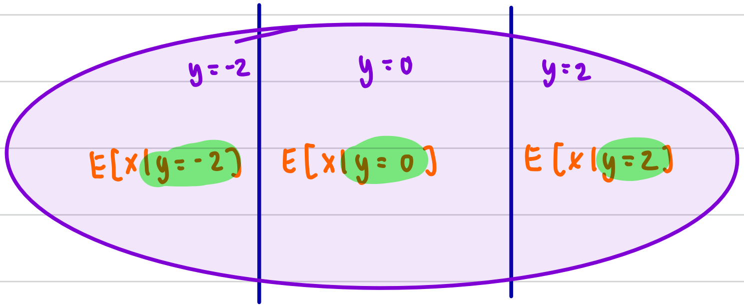
find $E(x+y)$

$$(1-2)\tfrac{1}{3} + (1+0)\,0 + (1+2)\,0 +$$
$$(2-2)\,0 + (2+0)\tfrac{1}{3} + (2+2)\tfrac{1}{6} +$$
$$(3-2)\,0 + (3+0)\tfrac{1}{6} + (3+2)\,0$$

For continuous r.v. ; $\int_y \int_x (x+y)\, f(x,y)\, dx\, dy$

Conditional expectation

$E[X \mid Y]$ = expectation of X for each group of Y



possible value of x

$1\, p(x=1|y) + 2\, p(x=2|y) + 3\, p(x=3|y)$

$X \backslash Y$	-2	0	2	$p(x)$
1	1/3	0	0	1/3
2	0	1/3	1/6	1/2
3	0	1/6	0	1/6
$p(y)$	1/3	1/2	1/6	1

$$E[x|y=0] = 1 \cdot p(x=1|y=0) \rightarrow \frac{p(x=1, y=0)}{p(y=0)} = 0$$

$$+\ 2 \cdot p(x=2|y=0) \rightarrow \frac{p(x=2, y=0)}{p(y=0)} = \frac{1/3}{1/2}$$

$$+\ 3 \cdot p(x=3|y=0) \rightarrow \frac{p(x=3, y=0)}{p(y=0)} = \frac{1/6}{1/2}$$

x can be only 2

$E[x|y=2] = 2$

because x can only be 2

constant

$Var[x|y=2] = 0$

$X\backslash Y$	-2	0	2	$p(x)$
1	1/3	0	0	1/3
2	0	1/3	1/6	1/2
3	0	1/6	0	1/6
$p(y)$	1/3	1/2	1/6	1

conditional prob $\quad P(x|y) = \dfrac{P(x,y)}{P(y)}$ $\quad$ joint prob / marginal prob

Random variable

- Discrete → pmf
- mixed
- continuous → pdf (density)

Distribution → $E[\;]$, $Var[\;]$, moment $[\;]$

marginal $\quad p(x), p(y) \rightarrow E[x]$

joint $\quad p(x,y) \rightarrow E[g(x,y)]$

now x is distributed within that group

conditional $\quad p(x|y) \rightarrow E[x|y]$

$= \dfrac{p(x,y)}{p(y)}$

use marginal probability

1) $E[x] = 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{2} + 3 \cdot \frac{1}{6} = \frac{11}{6}$

2) $E[x | y=-2] = 1 \cdot P[x=1 | y=-2] + 2 \cdot P[x=2 | y=-2] + 3 \cdot P[x=3 | y=-2]$

$P[x=2|y=-2] = \dfrac{P[x=2, y=-2]}{P(y=-2)}$

x can be only be 1 $\quad = 1 \cdot \dfrac{1/3}{1/3} + 0 + 0 = 1$

3) $E[x | y=0] = 1 \cdot P[x=1 | y=0] + 2 \cdot P[x=2 | y=0] + 3 \cdot P[x=3 | y=0]$

$= 0 + 2 \cdot \dfrac{1/3}{1/2} + 3 \cdot \dfrac{1/6}{1/2}$

$= 0 + \frac{4}{3} + 1 = \frac{7}{3}$

4) $E[x | y=2] = 1 \cdot p[x=1 | y=2] + 2 \cdot P[x=2 | y=2] + 3 \cdot P[x=3 | y=2]$

x can only be 2 $\quad = 0 + 2 \cdot \dfrac{1/6}{1/6} + 0 = 2$

5) $E[y | x=2] = -2 \cdot P[y=-2 | x=2] + 0 \cdot P[y=0 | x=2] + 2 \cdot P[y=2 | x=2]$

$= 0 + 0 + 2 \cdot \dfrac{1/6}{1/2} = \dfrac{2}{3}$

$X\backslash Y$	-2	0	2	$p(x)$
1	1/3	0	0	1/3
2	0	1/3	1/6	1/2
3	0	1/6	0	1/6
$p(y)$	1/3	1/2	1/6	1

✓ 6) $E[X^2Y] = \overset{1^2(-2)}{-2}\cdot\frac{1}{3} + \overset{1^2(0)}{0}\cdot 0 + \overset{1^2(2)}{2}\cdot 0 + \overset{2^2(-2)}{-8}\cdot 0 + \overset{2^2(0)}{0}\cdot\frac{1}{3} + \overset{2^2(2)}{8}\cdot\frac{1}{6} + \overset{3^2(-2)}{-18}\cdot 0 + \overset{3^2(0)}{0}\cdot\frac{1}{6} + \overset{3^2(2)}{18}\cdot 0 = -\frac{2}{3} + \frac{8}{6} = \frac{2}{3}$

✓ 7) $E[X^2Y \mid Y=0] = 0$

✓ 8) $E[X^2Y \mid Y=2] = 2\cdot\frac{0}{1/6} + 8\cdot\frac{1/6}{1/6} + 18\cdot\frac{0}{1/6} = 8$

$= E[2X^2 \mid Y=2]$
$= 2E[X^2 \mid Y=2] \rightarrow$ X can only be 2
$= 2\cdot 2^2\cdot\frac{1/6}{1/6} = 8$

✓ 9) $Var[X] = E[(X - \overset{11/6}{EX})^2] = \left(1-\frac{11}{6}\right)^2\cdot\frac{1}{3} + \left(2-\frac{11}{6}\right)^2\cdot\frac{1}{2} + \left(3-\frac{11}{6}\right)^2\cdot\frac{1}{6}$

or $E[X^2] - \left(\frac{11}{6}\right)^2$

$= \frac{25}{36}\cdot\frac{1}{3} + \frac{1}{36}\cdot\frac{1}{2} + \frac{49}{36}\cdot\frac{1}{6} = \frac{25}{108} + \frac{1}{72} + \frac{49}{216} = \frac{50}{216} + \frac{3}{216} + \frac{49}{216} = \frac{102}{216} = \frac{17}{36}$

✓ 10) $Var[X \mid Y=0] = E[(X - \overset{7/3}{E[X \mid Y=0]})^2 \mid Y=0]$

$Var[Y|X] = E[(Y - E[Y|X])^2|X]$

$= \left(1-\frac{7}{3}\right)^2\cdot\frac{0}{1/2} + \left(2-\frac{7}{3}\right)^2\cdot\frac{1/3}{1/2} + \left(3-\frac{7}{3}\right)^2\cdot\frac{1/6}{1/2}$

$= 0 + \frac{1}{9}\cdot\frac{2}{3} + \frac{4}{9}\cdot\frac{1}{3} = \frac{2}{27} + \frac{4}{27} = \frac{6}{27} = \frac{2}{9}$

or
$= E[X^2 \mid Y=0] - (E[X \mid Y=0])^2$

$= E[(X - E[X \mid Y=0])^2 \mid Y=0]$

✓ 11) $E[\underbrace{E[X|Y]}_{\text{function of } y}] = $

$\| \;\; E[X]$

E[X|Y] changes as y changes

$y=-2$	$y=0$	$y=2$
$E[X \mid Y=-2]$	$E[X \mid Y=0]$	$E[X \mid Y=2]$

$= 1\cdot P(Y=-2) + \frac{7}{3}\cdot P(Y=0) + 2\cdot P(Y=2)$

weighted by marginal prob of y

$= 1\cdot\frac{1}{3} + \frac{7}{3}\cdot\frac{1}{2} + 2\cdot\frac{1}{6} = \frac{1}{3} + \frac{7}{6} + \frac{2}{6} = \frac{11}{6}$ ✓ $= E[X]$

- If each component of a random vector (X, Y) takes on a continuum of values, then (X, Y) is a continuous random vector. Its joint distribution is completely described by the its PDF $f(x, y)$, satisfying the following properties:

  (i) $f(x, y) \geq 0$ for all (x, y)

  (ii) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$

  (iii) $P\{(X, Y) \in A\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I\{(x, y) \in A\} f(x, y) dx dy$

- There are other types of distributions, e.g. if X is discrete, such as age measured in years, but Y is continuous, such as wages.
- A function $g(X, Y): \mathbb{R}^2 \rightarrow \mathbb{R}$ is also a random variable. E.g. $X + Y, X^2 Y$

- **Marginal Distribution** of X is just another name for the distribution of X, but when used in context of a random vector (X, Y).
- It is used to emphasize the different between the distribution of X and the joint distribution of (X, Y).
- We can compute the marginal distribution of X from the joint distribution of (X, Y):

$$P\{X = x_i\} = \Sigma_{j=1}^{l} P\{X = x_i, Y = y_j\},$$

$$\text{or } f(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

- E.g. from the example above, the marginal probability $P\{X = 2\}$ is

$$P\{X = 2, Y = -2\} + P\{X = 2, Y = 0\} + P\{X = 2, Y = 2\} = \frac{1}{2}$$

- **Conditional Distribution** of X given $Y = y$, denoted by $P\{X|Y = y\}$, is the distribution of X when $Y = \text{y}$.
- The conditional probability $P\{X = x|Y = y\}$ is computed by

$$P\{X = x|Y = y\} = \frac{P\{X = x, Y = y\}}{P\{Y = y\}}$$

- Eg.: from the example above, the conditional probability

$$P\{X = 2|Y = 0\} = \frac{P\{X = 2, Y = 0\}}{P\{Y = 0\}} = \frac{1/3}{\left(0 + \frac{1}{3} + \frac{1}{6}\right)}$$

- Note that $P\{X = x|Y = y\}$ is a number, but $P\{X = x|Y\}$ is a function of Y, because the value of $P\{X = x|Y\}$ depends on the value of Y.
- You may think of the conditional distribution $P\{X|Y\}$ as the distribution of X within each group of Y.

MSF Chula

Expectation Revisit

- The expected value of a function of a random vector is computed in a usual way: summation or integration of each possible value times its probability or density.
- If involving just one variable, use marginal distribution:

$$E[g(X)] = \sum_{i=1}^{k} g(x_i)p(x_i) \text{ or } \int_{-\infty}^{\infty} g(x)f(x)dx$$

$$E[g(Y)] = \sum_{i=1}^{k} g(y_i)p(y_i) \text{ or } \int_{-\infty}^{\infty} g(y)f(y)dy$$

- If involving more than one variable, use joint distribution

$$E[g(X,Y)] = \sum_{i=1}^{k}\sum_{j=1}^{l} g(x_i, y_i)p(x_i, y_j), \text{ or}$$

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy$$

- **Conditional Expectation** of Y given X, denoted by $\boldsymbol{E[Y|X]}$ is the expectation of the conditional distribution of Y given X.
- Intuitively, if we divide the whole population into groups based on the value of X, then $E[Y|X]$ is the expected value of Y for each of the group.
- If X takes on a particular value x, we denote $E[Y|X = x]$.
- For example, consider the population of mutual funds. If $X = 1$ if a randomly selected mutual fund is equity fund, and $X = 0$ for other types of fund. Let R be uncertain return on a mutual fund. Then, $E[R|X = 1]$ is the expected return of equity funds and $E[R|X = 1]$ is the expected return of the other type of funds.
- The conditional expectation can be computed as:

$$E[\boldsymbol{Y}|X = x_i] = \sum_{j=1}^{l} \boldsymbol{y}_j P\{\boldsymbol{Y} = \boldsymbol{y}_j|X = x_i\}$$

- **Conditional Variance** of Y given X, denoted by $\boldsymbol{Var[Y|X]}$, is defined as

$$Var[Y|X] = E[(Y - E[Y|X])^2|X]$$

- Intuitively, it is the variance of Y for each group of population classified by the value of X.
- Since variance is a conditional expectation of a function of random variable, we can compute it in the similar way of finding conditional mean.
- Similar to the conditional expectation, $Var[Y|X]$ varies with X; so, it is random as well as $E[Y|X]$, but once the value of X is specified, $Var[Y|X = x]$ and $E[Y|X = x]$ are constant.

If not seperate → you need to use joint prob

Theorem: $E[g(X) + h(Y)] = E[g(X)] + E[h(y)]$

we know the value of x → can treat x as constant

Theorem: $E[g(X) + h(X)Y|X] = g(X) + h(X)E[Y|X]$

- Recall than $E[a + bX] = a + bE[X]$.
- So, intuitively, this theorem says that we can treat any function of X in $E[.|X]$ as a constant.

Theorem: $Var[\cancel{g(X)} + h(X)Y|X] = [h(x)]^2 Var[Y|X]$

$Var[\cancel{a} + bx] = b^2 Var[x]$

Law of Iterated Expectation

Theorem: $E[Y] = E\big[E[Y|X]\big]$

- Intuitively, this theorem say that we can calculation the mean of Y by two steps:
  - Firstly, divide the population into groups by the value of X and find the expectation of Y for each of the groups
  - Secondly, find expected value of the means from the first step.
- For example, if we want to know the mean income of Thai people, we can compute the (weighted) average of provincial per-capita incomes instead of collecting income of each Thai individuals to compute the average.

$E[E[y|x]] = E[y]$

$E[E[x|y]] = E[x]$

① find mean of each subgroup → then average again

② weighted by prob of that group

e.g. want to find $E[\text{income}]$ of the whole country

find $E[E[\text{income} | \text{province}]]$

BKK	CHMH	CH	Phuket
✓	✓	✓	✓
30k	20k	40k	25 k

$30 \cdot p(BKK) + 20\, p(CM) + 40 \cdot p(CH) + 25\, p(PH)$

$p(BKK) = \frac{20}{100}$

20 → # population in BKK

100 → # population in Thailand

GRADE :

MALE → 4 4 4 3 3

FEMALE → 4 4 2 2 3 3 3

$$E[\text{GRADE}] = \frac{4+4+4+3+3+4+4+2+2+3+3+3}{12}$$

$$= \frac{5}{5}\left(\frac{4+4+4+3+3}{12}\right) + \left(\frac{4+4+2+2+3+3+3}{12}\right)\frac{7}{7}$$

$$\underbrace{\left(\frac{5}{12}\right)}_{\text{prob of male in this room}}\left(\frac{4+4+4+3+3}{5}\right) + \frac{7}{12}\left(\frac{4+4+2+2+3+3+3}{12}\right)$$

Conditional

$$p(y \mid x) \neq p(y)$$

$|y$

interest in y, known x

No Relationship btw x & y

1) Independence $\Rightarrow$ $p(x,y) = p(x)\cdot p(y)$

stronger no relationship

$p(x \mid y) = p(x)$

$p(y \mid x) = p(y)$

If x & y are independent then $E[y \mid x] = E[y]$ and $E[x \mid y] = E[x]$

$\Downarrow$

2) Mean independence $\Rightarrow$ $E[y \mid x] = E[y]$

$E[x \mid y] = E[x]$

3) Uncorrelated

CHULALONGKORN BUSINESS SCHOOL FLAGSHIP FOR LIFE

Independence

- Two random variables X and Y are independent if, for any set A, B

$$\underbrace{P\{X \in A, Y \in B\}}_{\text{joint prob}} = \underbrace{P\{X \in A\} \cdot P\{Y \in B\}}_{\text{multiplication of marginal prob}}$$

- This means that knowing something about X reveals nothing about Y and vice versa
- If X is independent of Y, then any function of X is also independent of Y
- For discrete random vector (X, Y), independence means for all *i, j*

$$P\{X = x_i, Y = y_j\} = P\{X = x_i\} \cdot P\{Y = y_j\}$$

- For a continuous random vector with PDF $f(x, y)$, independence means

$$f(x, y) = f(x) \cdot f(y)$$

$$P(x|y) = \frac{P(x,y)}{P(y)} = \frac{P(x)\cdot \cancel{P(y)}}{\cancel{P(y)}} = P(x)$$

e.g. sample 1 stock in mkt ⇒ prob it would be positive cannot know it from knowing other stock
don't know anything and sample from population
know x doesn't make you know y

$$P(y|x) = P(y)$$

x & y are independent

$$E[y|x] = \sum y_i\, P(y_i|x) = \sum y_i\, P(y_i) = E[y]$$

↳ is the same as weighted by marginal prob

MSF Chula

CHULALONGKORN BUSINESS SCHOOL
FLAGSHIP FOR LIFE

AACSB ACCREDITED · EFMD EQUIS ACCREDITED · EFMD EPAS ACCREDITED

- The independence implies that conditional is the same as marginal probability or distribution.
- For example,

$$P\{Y = y_j | X = x_i\} = \frac{P\{X = x_i, Y = y_j\}}{P\{X = x_i\}} = \frac{P\{X = x_i\}P\{Y = y_j\}}{P\{X = x_i\}} = P\{Y = y_j\}$$

- Similarly for the continuous case, $f(y|x) = \frac{f(x,y)}{f(x)} = \frac{f(x)f(y)}{f(x)} = f(y)$
- $P\{Y|X\} = P\{Y\}$ means that knowing the value of X cannot tell more information about the probability of Y.
- Similarly, independence implies $P\{X|Y\} = P\{X\}$; so, knowing the value of Y also tells nothing new about X.
- Independence is not a transitive property

x independent of y
y " ——— " z
x & z are not necessary be independent
e.g. x = z

Mean Independence

- Recall that the value of the conditional mean $E[Y|X]$ generally varies with X; thus, it is a random variable.
- **Y is mean independent of X**, if $E[Y|X]$ does not vary with X.
- In other words, Y is mean independent of X, if $E[Y|X]$ is the same for all groups of X.
- E.g. Let W be wage and S be an indicator variable for male. $E[W|S = 1]$ is the expected value of the wages of males, and $E[W|S = 0]$ is the expected wage of females, both of which are constant. However, $E[W|S]$ is generally a function of the gender variable S.
- if $E[W|S = 1] = E[W|S = 0]$, then $E[W|S]$ doesn't depend on S. In this case, wage is mean independent of gender identity.
- Remark: That Y is mean independent of X doesn't imply that X is mean independent of Y.

y is mean independent of x

mean of y doesn't depend on x

$\underbrace{E[y \mid x]}_{\text{function of } x}$

c	c	c

no matter group of x we are in
y always equal to c

$E[y|x] = c = E[y]$

mean of popu = average mean of 3 group = c

!!! x is not necessary to be mean independent of y

MSF Chula

CHULALONGKORN BUSINESS SCHOOL
FLAGSHIP FOR LIFE

$Var[X] = E[(X-E(X))^2] = E[X^2] - (E(X))^2$

Covariance and Correlation

- **Covariance** between X and Y, denoted by $Cov[X,Y]$ or σ_{XY}, is defined as

on average

$$Cov[X,Y] = \boldsymbol{E}\left[\left(\boldsymbol{X}-\boldsymbol{E}(\boldsymbol{X})\right)\left(\boldsymbol{Y}-\boldsymbol{E}(\boldsymbol{Y})\right)\right] = E[XY] - E[X]E[Y]$$

+ + move together in the same direction

− move in opposite direction

- It measures the extent to which one random variable deviates from its mean relative to how the other random variable moves from its mean.

- Units of the covariance is the unit of X times the unit of Y. So, it is useful to “normalize” to become unit-less

- **Correlation** between X and Y is $\boldsymbol{corr[X,Y] = \frac{Cov[X,Y]}{\sigma_X \sigma_Y}}$ always positive

- We say that **X and Y are uncorrelated if $corr[X,Y] = 0$**

covariance

weight

- age $\Rightarrow$ kg · year 50
- height $\Rightarrow$ kg · cm 500 cannot say that this is better

Properties of covariance and correlation

- For any random vector (X, Y), $-1 \leq Corr[X, Y] \leq 1$
- For any random vector (X, Y, W) and constants a, b

$\mathrm{cov}(x, a) = 0$

$$Cov[(\cancel{a} + bX), Y] = \underline{b}Cov(X, Y)$$

$$Cov[(X + W), Y] = Cov(X, Y) + Cov(W, Y)$$

- For any random vector (X, Y) and constants a, b

$$Var[a\mathbf{X} + b\mathbf{Y}] = a^2 Var(\mathbf{X}) + b^2 Var(\mathbf{Y}) + 2ab Cov(\mathbf{X}, \mathbf{Y})$$

Properties

$E[c] = c$

$E[a+bX] = a + bE[X]$

$E[X+Y] = E[X] + E[Y]$

$E[XY] = E[X] \cdot E[Y] + \underline{cov(X,Y)}$

$cov(X,Y) = E[XY] - E[X] \cdot E[Y]$

$E[XY] = \underline{cov(X,Y)} + E[X] \cdot E[Y]$

$Var[c] = 0$

$Var[a+bX] = b^2 \, var[X]$

$Var[X+Y] = Var[X] + Var[Y] + 2cov(X,Y)$

$E[\{(X+Y) - E(X+Y)\}^2]$

$X + Y - E[X] - E[Y]$

$(X - E[X]) + (Y - E[Y])$

$E[(A+B)^2] = E[A^2] + E[B^2] + 2E(AB)$

strongest

Independence

imply

$p(x|y) = p(x)$

$p(y|x) = p(y)$

Mean Independence

imply

Uncorrelated

$E[XY] = E[X]E[Y] + \underbrace{cov(X,Y)}_{\to 0}$

Prove that mean independent → uncorrelated

$cov(x,y) = 0$

assume $E[y|x] = c = E[y]$

definition of covariance

$E[XY] = E[X]E[Y]$

prove that
If you have mean independent
then you have uncorrelated

can also write $E[E[xy|y]]$

x is treated as constant

$E[xy] = E[E[xy|x]]$; law of iterated expectation (LIE)

$= E[x\,\underbrace{E[y|x]}_{c}]$

$= E[x \cdot c]$

$= E[x] \cdot \overset{E[y]}{c}$

$= E[x]E[y]$

Outline

Population

Sample
(Data, observations)

Statistical Review

- Random Sampling
- Estimation
- Finite-sample Properties of Estimators
  - ☆ Biasness   good estimator should be unbiased
  - Variance and Efficiency
- Large-sample Properties of Estimators
  - ☆ Consistency   good estimator should be consistent
  - Weak Law of Large Number
  - Continuous Mapping Theorem
  - Central Limit Theorem

Random Sampling

possible value must be the same as we draw r.v. from the same population
has exactly the same distribution as other x

- If $X_1, X_2, \ldots, X_n$ are independent random variables with the same distribution X, the we say that $X_1, X_2, \ldots, X_n$ are **Independent and Identically Distributed (iid)** sample of size n.

random variable
sample n times
they're independent
know x_1 doesn't know x_2

- Notation: $\boldsymbol{X_1, \ldots, X_n \sim iid\ X}$

time that we sample from population

- We want to learn about features of the distribution of X from this random sample of data. Each feature that we're interested in learning about is called a **parameter**. We generally use a Greek letter to represent a parameter.
- For example, parameters are the mean (μ), variance (σ^2), ... etc. of the distribution of X.
- Typically, we don't know the distribution of X, as we cannot observe all the **population** of X. E.g. X may be the income of investors.

we don't know distribution

Population

PARAMETER

- interested
- unknown
- fixed and constant

$E[X]$, μ , θ, β, α greek letter

$Var[X]$, σ^2

Sample

(Data, observations)

ESTIMATOR → is random depend on data we use

method / rule / function of (observed) data

⇓ plug in data to estimator then get estimate

estimate

MSF Chula

CHULALONGKORN BUSINESS SCHOOL FLAGSHIP FOR LIFE

AACSB ACCREDITED · EFMD EQUIS ACCREDITED · EFMD EPAS ACCREDITED

Estimation

- If the distribution of X is unknown, a parameter θ of the distribution of X is unknown as well. We want to provide a "best guess" or an **estimate** for the unknown value of θ based on the random sample of size n. estimator $\hat{}$ hat $\tilde{}$ tilde $\bar{}$ bar

parameter

- **An estimator for θ** is a rule for using the random sample to construct an estimate. In other words, an estimator is a function of random variables.
- Notation: $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$
- Note **that a parameter, θ, is unknown and non-random**, whereas **an estimator, $\hat{\theta}_n$, is random and has a distribution**.
- For a given parameter, there can be several estimators.

- E.g. Let $X_1, \dots, X_n \sim iid\ X$, and $\theta = E[X]$ denotes the parameter of interest. A natural estimator of θ is the **sample mean**:

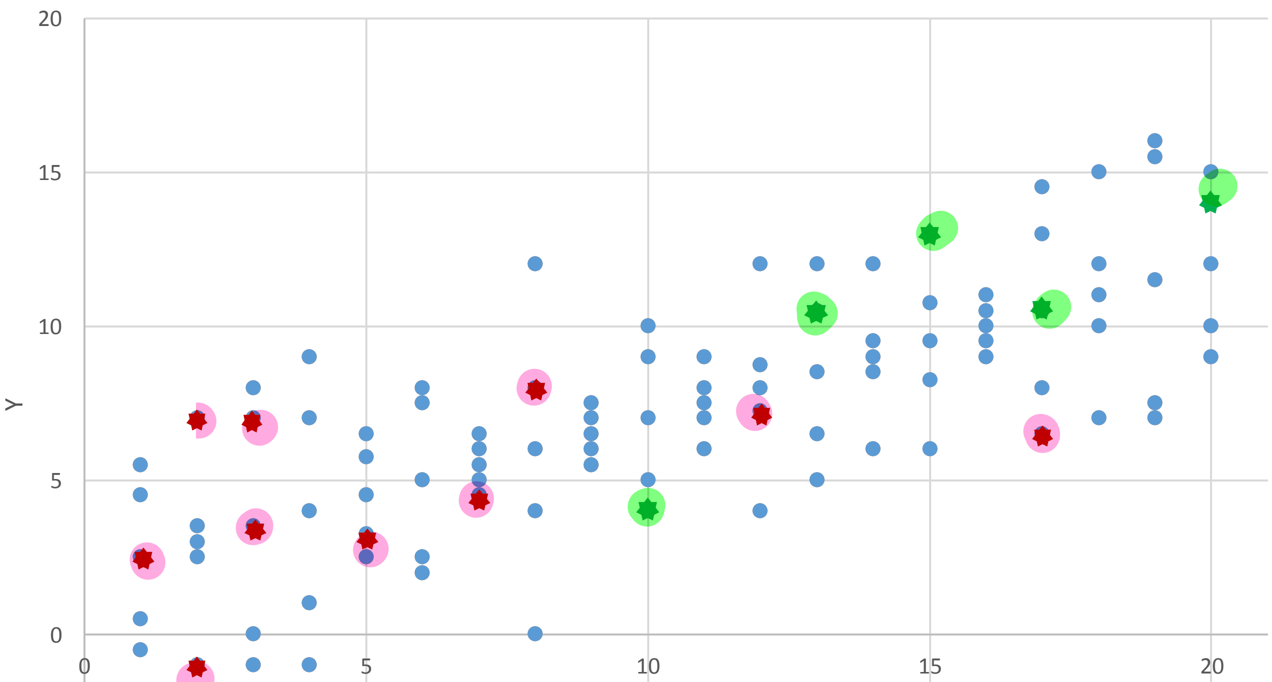
$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \equiv \bar{X}_n$$

- E.g. We can also have $\tilde{\theta}_n = (X_1 + X_n)/2$ as an estimator for $E[X]$.

- E.g. Let $X_1, \dots, X_n \sim iid\ X$, and $\theta = Var[X]$ denotes the parameter of interest. A natural estimator of θ is the **sample variance**:

$$\hat{\theta}_n = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

Population's Parameter vs Estimators and Sample's Estimates



belongs to population

Parameter

Expected Value of X:

$E[X] = \mu_X = 10$

rule / function

Estimators for $E[X]$

$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

1st observation

last observation

$\tilde{X}_n = \frac{X_1 + X_n}{2}$

x tilde

plug in different data in estimator we get different estimates

estimator

estimate

$\bar{X}_{10} = 6$ $\quad$ $\bar{X}_{15} = 9$ $\quad$ $\bar{X}_5 = 15$

estimate

$\tilde{X}_{10} = 9$ $\quad$ $\tilde{X}_{15} = 10.5$ $\quad$ $\tilde{X}_5 = 15$

expectation belongs to population

1 parameter $E[X]$ or μ_X

sample mean!!!

$\bar{X}_n$ $\tilde{X}_n$ we can have many estimators

estimator

<u>random</u> not constant!!!

can use any of two

GOOD ESTIMATOR ?

① Unbiased $\not\Leftrightarrow$ not imply to each other ② Consistent

↳ true on average

↳ E[estimator] = parameter

↳ converge to

become very close to ↳ true value (which is the parameter)

when N is large

↳ number of observations

plim [estimator] = parameter ⇒ much easier to prove than unbiase

Finite-Sample Properties

- These are properties of the distribution of an estimator that hold regardless of sample sizes.

Biasedness is defined as the difference between the expected value of the estimator and the parameter of interest:

$$\boldsymbol{Bias(\hat{\theta}_n) = E[\hat{\theta}_n] - \theta}$$

can find expectation

has distribution

random

parameter

we don't have to know

- An estimator is **unbiased** if $Bias(\hat{\theta}_n) = 0$, that is, $\boldsymbol{E[\hat{\theta}_n] = \theta}$

, we can prove it with statistics

- An estimator is biased upwards or positive biased if $E[\hat{\theta}_n] > \theta$
- An estimator is biased downwards or negative biased if $E[\hat{\theta}_n] < \theta$

Parameter : μ_X or $E[X]$

Sample : $X_1, X_2, X_3, \ldots, X_n \overset{iid}{\sim} X$ ← chance that X_i can happen is the same as chance that X can happen

estimator : $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

constant

$$E[\bar{X}_n] = E\left[\frac{1}{n}(X_1 + X_2 + \ldots + X_n)\right]$$

$$= \frac{1}{n}\left(\underbrace{E[X_1]}_{E[X]} + \underbrace{E[X_2]}_{E[X]} + \ldots + \underbrace{E[X_n]}_{\mu_X}\right)$$

$$= \frac{1}{\cancel{n}}(\cancel{n}\mu_X)$$

$$= \mu_X$$

$\therefore \bar{X}_n$ is unbiased estimator

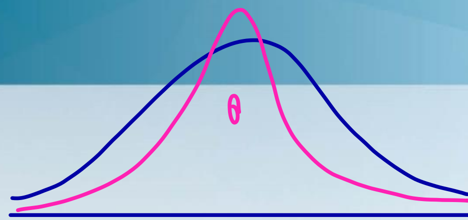
- Eg.: Let $X_1, \dots, X_n \sim iid\ X$. The **sample mean $\bar{X}_n$** is an unbiased estimator of $E[X]$ because

$$E[\bar{X}_n] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n} E[X] = E[X]$$

- Eg.: Let $X_1, \dots, X_n \sim iid\ X$, and $\boldsymbol{\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}$. It can be showed that $\hat{\sigma}_n^2$ is an unbiased estimator of $Var[X]$

- Then, the estimator $\hat{s}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ is a biased estimator of the variance of X because

$$E[\hat{s}_n^2] = E\left[\left(\frac{n-1}{n}\right)\hat{\sigma}_n^2\right] = \left(\frac{n-1}{n}\right)Var[X]$$



Efficiency of Estimator: the distribution of the estimator to be concentrated as much as possible around the underlying parameter θ.

This leads to more than one definition of "efficiency," but one idea is to pick an estimator with **the smallest variance.** proxy

- E.g. Let $X_1, \dots, X_n \sim iid\ X$. The variance of $\bar{X}_n$ is

$$Var[\bar{X}_n] = Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2} Var\left[\sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} Var[X_i] = \frac{1}{n^2}\sum_{i=1}^{n} Var[X] = \frac{1}{n} Var[X]$$

- E.g. Let $X_1, \dots, X_n \sim iid\ X$ and $\hat{\theta}_n = X_n$ be an estimator for $E[X]$. Then, $\hat{\theta}_n$ is another unbiased estimator, but it is less efficient than $\bar{X}_n$ since

$$Var\left[\hat{\theta}_n\right] = Var[X_n] = Var[X] > \frac{1}{n} Var[X] = Var[\bar{X}_n]$$

$\bar{X}_n \Rightarrow E[X]$

unbiased

$$E[\tilde{X}_n] = E\left[\frac{X_1 + X_n}{2}\right] = \frac{1}{2}E[X_1 + X_n]$$

$$= \frac{1}{2}\left[\underbrace{E[X_1]}_{\mu} + \underbrace{E[X_n]}_{\mu}\right]$$

$$= \mu$$

$$Var[\tilde{X}_n] = Var\left[\frac{X_1 + X_n}{2}\right]$$

$$= \frac{1}{4} var[X_1 + X_n]$$

$$= \frac{1}{4}\left(\underbrace{var(X_1)}_{var(X)} + \underbrace{var(X_n)}_{var(X)} + \cancel{2\,cov(X_1, X_n)}^{0}\right)$$

$$= \frac{1}{2} var(X)$$

iid $\Rightarrow$ X_1 & X_n are independent

$\Downarrow$

X_1 is mean independent of X_n

X_n " " X_1

$\Downarrow$

uncorrelated

$$Var[\bar{X}_n] = var\left[\frac{1}{n}(X_1 + X_2 + \dots + X_n)\right]$$

$$= \frac{1}{n^2}[\underbrace{\cancel{var(X_1)}}_{\sigma_X^2} + \underbrace{\cancel{var(X_2)}}_{\sigma_X^2} + \dots + \underbrace{\cancel{Var(X_n)}}_{\sigma_X^2}]$$

no need to care about cov

$$= \frac{n}{n^2}\sigma_X^2$$

$$= \frac{1}{n} var(X)$$

$\Rightarrow$ more observations $\rightarrow$ more precise estimate / closer estimate to parameter

$\bar{X}_n$ is better than $\tilde{X}_n$

require large N

Large-Sample Properties

- These are properties of the distribution of an estimator that hold *approximately* for large sample sizes.

Consistency: an estimator $\hat{\theta}_n$ of a parameter θ is consistent if the estimator converges in probability to θ. true parameter

probability limit

- Notation: $\hat{\theta}_n \xrightarrow{P} \theta$ or $\text{plim}(\hat{\theta}_n) = \theta$
- Roughly speaking, **when the sample size *n* gets large, $\hat{\theta}_n$ gets close to θ with high probability**. Imagine the distribution of $\hat{\theta}_n$ collapses roughly to a single point θ.

everything in sequence is constant

Sequence $\{\frac{1}{n}\}$: $\overbrace{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \ldots, \frac{1}{n}}, \ldots$

approach to zero

$\lim_{n\to\infty} \{\frac{1}{n}\} = 0$

of estimator

Sequence $\{\hat{\theta}_n\}$ or sequence of $\{\bar{x}_n\}$

↳ tell the sample size, how many observations we use to estimate

everything is random

$\overbrace{\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4, \bar{x}_5, \bar{x}_6, \ldots, \bar{x}_n}, \ldots$

↳ most value of $\bar{x}_n$ is 100% close to...

probability limit $\operatorname{plim}_{n\to\infty} \{\bar{x}_n\}$: tell the single point that this random sequence approach to

↓

if it equal to parameter

↓

it is consistent

The chance that we get positive returns

$$X \begin{cases} 1 & + \text{ return} \\ 0 & \text{otherwise} \end{cases}$$

parameter of interest

$X \sim \text{Bernoulli}(p)$ p is also expectation of X ; $E[X] = 1 \cdot p + \underbrace{0 \cdot (1-p)}_{0}$

Estimator : $\bar{X}_n$

Suppose $p = 0.75$; $X \sim \text{Bernoulli}(0.75)$

n (times that sample)	$\bar{X}_n$ (possible value)
1	$\{0, 1\}$ ($\frac{1}{4}$, $\frac{3}{4}$)
2	$\{0, 1, \frac{1}{2}\}$ ($\frac{0+0}{2}$, $\frac{1+1}{2}$, $\frac{1+0}{2}$)
3	$\{0, \frac{1}{3}, \frac{2}{3}, 1\}$
4	$\{0, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, 1\}$

distribution (pmf) of $\bar{X}_n$

$p(\bar{x}_1 = 0) = p(x_1 = 0) = \frac{1}{4}$ or 0.25

$p(\bar{x}_1 = 1) = p(x_1 = 1) = \frac{3}{4}$ or 0.75

$p(\bar{x}_1 = 0) = p(x_1 = 0, x_2 = 0)$; x_1 & x_2 are independent

$= p(x_1 = 0)\, p(x_2 = 0)$

$= \frac{1}{4} \cdot \frac{1}{4}$

$p(\bar{x}_2 = 1) = p(x_1 = 1 \text{ and } x_2 = 1)$

$= p(x_1 = 1)\, p(x_2 = 1)$

$= \frac{3}{4} \cdot \frac{3}{4}$

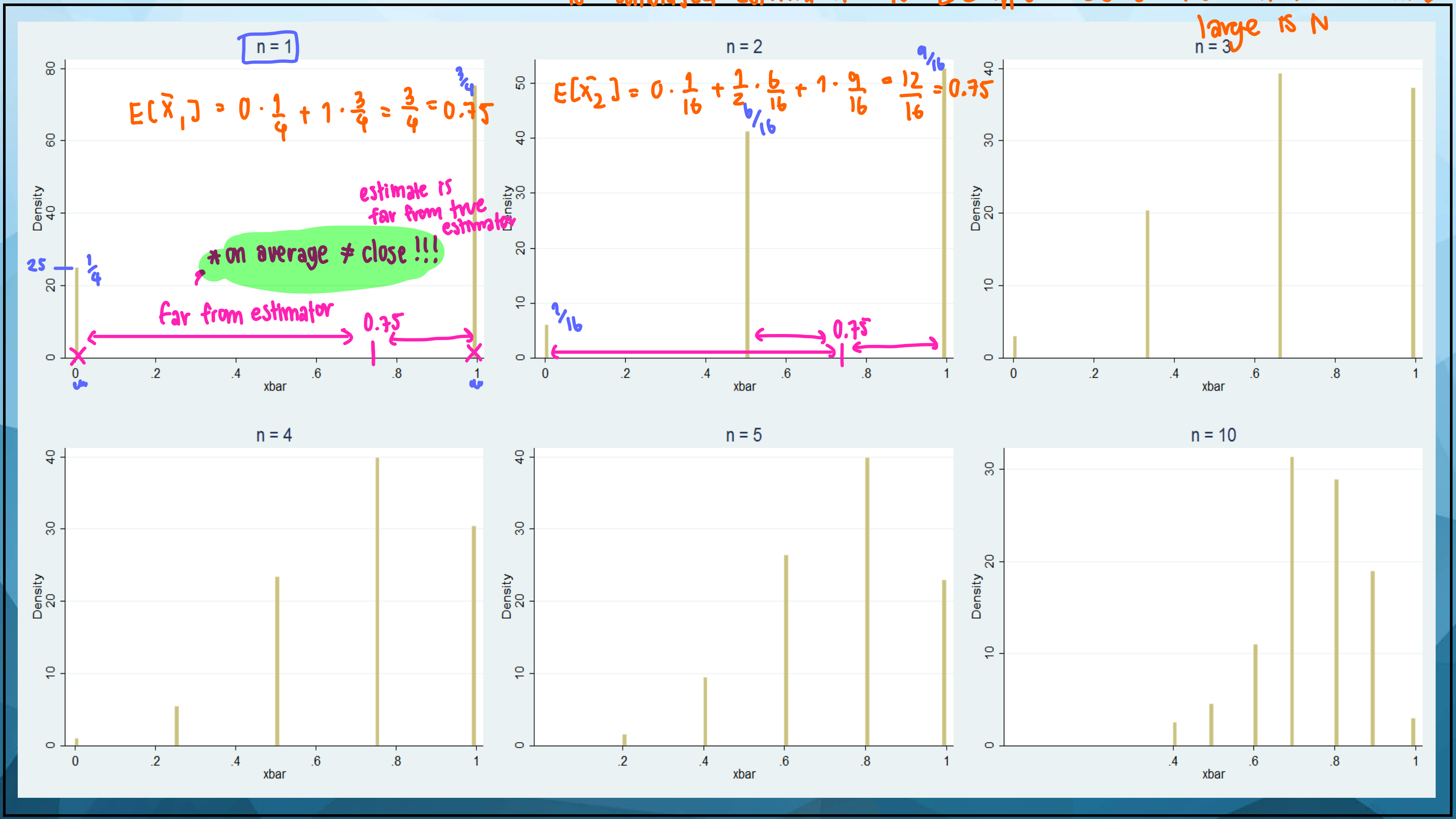
MSF Chula

CHULALONGKORN BUSINESS SCHOOL
FLAGSHIP FOR LIFE

as n change, the distribution change

Distributions of $\overline{X}_n$, for $X_1, \ldots, X_n \sim iid$ Bernoulli(0.75)

↳ is unbiased estimator as $E[\bar{X}_n] = E[X] = 0.75$ no matter how large is N on average

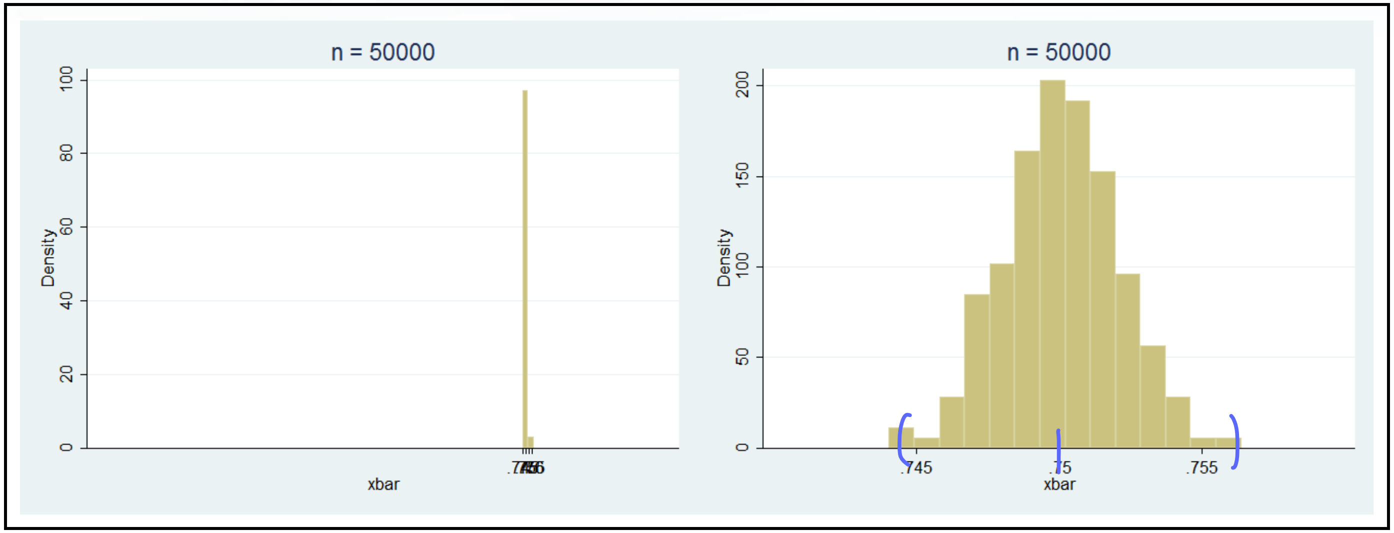



n = 10
n = 50
n = 100
n = 200
n = 500
n = 1000
n = 5000
n = 10000
n = 50000
Density
xbar
parameter
0.75
0.75
0.75
0.75
approach to parameter
<random very close
to true parameter>
consistency → corrupt to one point
almost
100%
equal to 0.75
&
converge in probability
to 0.75

- Note: a consistent estimator $\hat{\theta}_n$ is still random, not exactly a constant θ, even when the sample size n is very large.

$\bar{X}$ still random → not exactly equal to 0.75 but very close to 0.75

- E.g. if we zoom in the distribution of $\bar{X}_{50,000}$ in previous example,



- it is still distributed around, but very close to, the true mean

$E[X] = 0.75$

Way to prove plim

I

Weak Law of Large Number Theorem

converge in prob (plim)

- Let $\boldsymbol{X_1, \ldots, X_n \sim iid\ X}$. Suppose $Var[X] < \infty$, then $\bar{\boldsymbol{X}}_{\boldsymbol{n}} \xrightarrow{P} \boldsymbol{E[X]}$

Continuous Mapping Theorem (CMT)

plim can expand in any thing if it's continuous function

- if $A_n, n \geq 1$ and $B_n, n \geq 1$ are sequences of random variables and a and b are constants such that

$$A_n \xrightarrow{P} a \text{ and } B_n \xrightarrow{P} b$$

and $g: \mathbb{R}^2 \to \mathbb{R}$ is continuous at (a, b), then $g(A_n, B_n) \xrightarrow{P} (a, b)$

- Roughly speaking, $plim(.)$, if existing, can be broken down in any continuous functions.

$\boxed{\mathrm{I}}$

Sample mean $\xrightarrow{p}$ true mean

$\Downarrow$

$$\operatorname{plim}\left(\frac{1}{n}\sum_{i=1}^{n} \boxed{iid}\right) \qquad \operatorname{plim}\left(\overline{\boxed{iid}}_n\right) = E\left[\boxed{iid}\right]$$

- E.g. Let $A_n \xrightarrow{P} a$ and $B_n \xrightarrow{P} b$ and $g(x, y) = x + y$. Then, by CMT

$$A_n + B_n = g(A_n, B_n) \xrightarrow{P} g(a, b) = a + b$$

- Similarly, the CMT implies
  - $plim(A_n + B_n) = plim(A_n) + plim(B_n)$
  - $plim(A_n - B_n) = plim(A_n) - plim(B_n)$
  - $plim(A_n \times B_n) = plim(A_n) \times plim(B_n)$
  - $plim(A_n \div B_n) = plim(A_n) \div plim(B_n)$ if $plim(B_n) \neq 0$

$E[A \cdot B] \neq E[A]E[B]$

multiplication is continuous function

$E\left[\frac{A}{B}\right] \neq \frac{E[A]}{E[B]}$

Example $X_1, X_2, \dots, X_n \overset{iid}{\sim} \text{Bernoulli}(\tfrac{3}{4})$

converge to mean of X from weak law of large number

can expand into continuous function

$$\text{plim}\left(2^{\frac{1}{n}\sum_{i=1}^{n} X_i}\right) \overset{\text{by CMT}}{=} 2^{\text{plim}\left(\frac{1}{n}\sum X_i\right)}$$

from CMT

$$\overset{\text{by WLLN}}{=} 2^{E[X]}$$

$$= 2^{3/4}$$

① Check which one is unbiased
② Check which one has lower variance

parameter : $E[X]$

estimator 1 $\Rightarrow$ $\tilde{X}_n = \frac{X_1 + X_n}{2}$

▸ Check unbiased estimator : $E[\tilde{X}_n] = E\left[\frac{X_1+X_n}{2}\right]$

$X_1, X_2, \ldots, X_n \overset{iid}{\sim} X$

$\frac{1}{2}[E(X_1)+E(X_n)]$

$= \frac{1}{2}[\mu+\mu] = \mu$

$= \frac{1}{2} E[X_1 + X_n]$

chance that X_i can happen is the same as chance that X can happen

$= \frac{1}{2}(2E[X])$

$= E[X]$ $\longrightarrow$ $\tilde{X}_n$ is unbiased estimator for $E[X]$

▸ Find variance : $Var[\tilde{X}_n] = Var\left[\frac{X_1+X_n}{2}\right]$

$= \frac{1}{4} Var[X_1+X_n]$

X_1 & X_n are independent

$= \frac{1}{4} \cdot (Var[X_1] + Var[X_n] + \underbrace{\cancel{cov(X_1, X_n)}}_{0})$

$= \frac{1}{2} \cdot Var[X]$

parameter : $E[X]$

estimator 2 $\rightarrow \hat{X}_n = \frac{3X_1 + 5X_n}{8}$

- Check unbiased estimator : $E[\hat{X}_n] = E\left[\frac{3X_1+5X_n}{8}\right]$

$X_1, X_2, \dots, X_n \overset{iid}{\sim} X$

$\frac{3}{8}E[X_1] + \frac{5}{8}E[X_n]$; $E[X_1] \to \mu$, $E[X_n] \to \mu$

$= \left(\frac{3+5}{8}\right)\mu$

$= \mu$

$= \frac{3}{8}E[X_1] + \frac{5}{8}E[X_n]$ ✓

chance that x_i can happen is the same as chance that x can happen

$= \frac{3}{8}E[X] + \frac{5}{8}E[X]$

$= E[X] \longrightarrow E[\hat{X}_n]$ is unbiased estimator for $E[X]$

- Find variance : $Var[\hat{X}_n] = Var\left[\frac{3x_1+5X_n}{8}\right]$

X_1 & X_n are independent

$\frac{9}{64}Var(x_1) + \frac{25}{64}Var(x_n)$

$= \frac{34}{64}Var(x) > \frac{32}{64}var(x)$

$= \frac{1}{64} Var[3X_1 + 5X_n]$

$25\,cov(\cancel{x_1}, X_n) \to 0$

$= \frac{1}{64} \cdot \left(9\underbrace{Var[X_1]}_{Var[X]} + 25\underbrace{Var[X_n]}_{Var[X]} + \underbrace{cov(3X_1, 5X_n)}_{25\,cov(x_1, X_n) \to 0}\right)$

$= \underbrace{\frac{34}{64}}_{0.531} \cdot Var[X]$

$\therefore \tilde{X}_n$ has lower variance than $\hat{X}_n$ ← less efficient

↳ better estimator for $E[X]$ ✓

Review :

No relationship btw X & Y :

- Independent $\quad p(x,y) = p(x)\,p(y)$ implies $p(y|x) = p(y)$ and $p(x|y) = p(x)$

  know x (y) doesn't mean we know y (x)

  same as population mean

  ⇓
- Mean Independent $\quad E[y|x] = c = E(y)$

  conditional expectation is the same as population expectation

  ⇓
- Uncorrelated

  $cov(x,y) = 0$

  $corr(x,y) = 0$

Population $\qquad$ **Sample**

(Data, observations)

PARAMETER ← ESTIMATOR

- unknown
- fixed / constant

  mean or average

  $E[X]$, μ

- function of observations
- random → has distribution

  sample mean $\bar{X}_n$, $\sim$, $\hat{\ }$ ← attached to parameter

Good Estimator

Unbiased $\Rightarrow$ on average = parameter

E [estimator]

$\not\Leftrightarrow$

Consistent $\Rightarrow$ converge

when we have large N, estimate is almost 100% close to true parameter plim [estimate]

Sampling Distribution

- Sampling distribution is the distribution of an estimator.
- It depends on the unknown distribution of the underlying random variables $X_1, \dots, X_n$ used with the estimator.
- Hence, it is generally difficult to compute, unless we are willing to make strong assumptions on the distribution of the random variables.
- E.g. Let $X_1, \dots, X_n \sim iid\ X$. Assume $X \sim N(\mu_x, \sigma_x^2)$ and $\bar{X}_n$ is the estimator. Then, since $\bar{X}_n$ is a linear combination of normal random variables, it is also distributed normally with the mean and variance as calculated before, i.e.

$$\bar{X}_n \sim N(\mu_x, \frac{\sigma_x^2}{n})$$

Central Limit Theorem (CLT)

Let $\boldsymbol{X_1, \dots, X_n \sim iid\ X}$. Suppose $0 < Var[X] < \infty$. Then

$$\frac{\bar{X}_n - \mu_X}{\frac{\sigma_X}{\sqrt{n}}} \xrightarrow{d} N(0,1)$$

- The key message of this theorem is that the sampling distribution of the sample mean $\bar{X}_n$ from any iid sample will become approximately Normal when sample size is sufficiently large.
- In other words, it suggests that, approximately for a large iid sample,

$$\bar{X}_n \sim N(\mu_x, \frac{\sigma_x^2}{n})$$

regardless of the underlying distribution of X.