Today's outline

Midterm

Hypothesis, classification

Presentation

Q 2

(B) 2 r.v.s X, Y

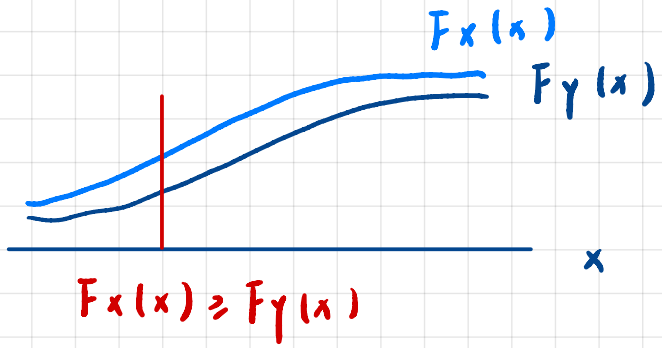
$P(X \le Y) = 1$

$F_X(x) \ge F_Y(x) \ \forall x$ T or F ?

$F_Y(x) = P(Y \le x)$

$\le P(X \le x)$

$\{Y \le x\} \subset \{X \le x\} = F_X(x)$ ⇒ True



?

(C) $X, Y, \max\{X, Y\}$

$F_{\max\{X,Y\}}(x) = F_X(x) \cdot F_Y(x)$, given $X \perp\!\!\!\perp Y$

$F(\max\{X,Y\} \mid x)$

$= P(\max\{X,Y\} \le x)$

$= P(\{X \le x\} \cap \{Y \le x\})$

$= P(X \le x) \cdot P(Y \le x)$

$F_X(x) \quad F_Y(x)$

?

Q3 (c)

$X \perp\!\!\!\perp Y$

X, Y jointly normal and $Cov(X,Y)=0$ True

$\Rightarrow E[Y|X] = E[X]$

(E) $X \perp\!\!\!\perp Y$, $Cov(X^2, Y) = 0$ True

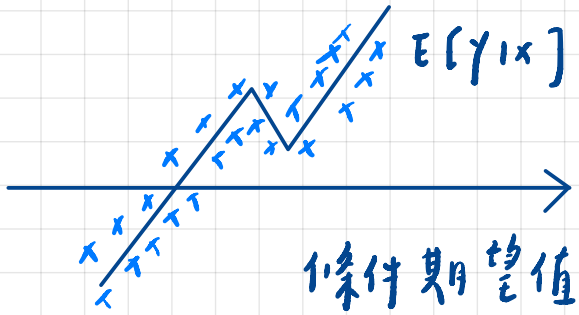
Q4 (B)

$E[Y|X] = aX + b$ OLS $\rightarrow E[Y|X]$ True

Ans: OLS : best linear predictor

$E[Y|X]$ best predictor

(E) $Cov(X,Y) > 0 \Rightarrow E[Y|X=x]$ increases in x False



條件期望值 > 0，可能非全遞增。

Q5 $\Rightarrow Var(x_1) > Var(x_2)$

(x_1, x_2) $a_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ $(x_2, 2x_1) \Rightarrow \tilde{a_1} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$?

$\Rightarrow Var(2x_1) > Var(x_2)$

a_1 = solution

$$\max_a Var(a^T x) \quad s.t. \quad a^T a = 1$$

Q6

$(x_1, 2x_2)$ $\tilde{a_2} = (\ )$?

$\tilde{a_2}^T \tilde{a_1} = 0 \Rightarrow \tilde{a_2} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

Q8

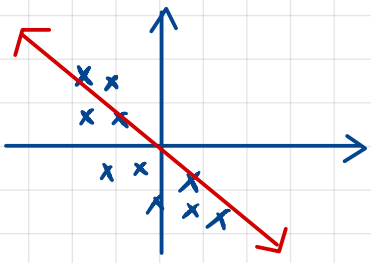
x_1, x_2, $a_1 = \begin{pmatrix} 3/5 \\ 4/5 \end{pmatrix}$ $Cov(x_1, x_2) \overset{>}{=} 0$?

$<$

$Var(a^T x) = Var(a_{11} x_1 + a_{12} x_2)$

$= Var(a_{11} x_1) + Var(a_{12} x_2) + 2 a_{11} a_{12} Cov(x_1, x_2)$

if $Cov(x_1, x_2) < 0 \Rightarrow Var(a_{12} x_2) < 0$ 才有 max Var



Q9. 10. 11

Find the pair of closest point for the d(x.y) given

Q13. 14. 15

Find ols $\hat{\beta}_1, \hat{\beta}_2$

for $y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon$ given data

Sol <1>

$\hat{\beta}_1, \hat{\beta}_2$ are sample means for $x_1 = 1$ & $x_2 = 1$

Sol <2>, <3>

$\hat{\beta}_1, \hat{\beta}_2$ are the solutions

$$\min_{\beta_1, \beta_2} 1/n \Sigma (y_i - \beta_1 x_{i1} - \beta_2 x_{i2})^2$$

solve by F.O.C or 配方法

↳ 拋物線最小值

Sol <4>

ols has closed form

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (X^T X)^{-1} X^T Y$$

Hypothesis Testing

H_0 : Statement 1

H_1 : Statement 2

真项只有一個, which one ?

we have data $X = x_1, x_2, \dots, x_n$

A test is data-driven decision rule $D(x)$

$D(x) \in$ { Reject H_0 ; Accept H_0 }

A good test : { Reject H_0 when H_0 false ; Accept H_0 when H_0 true }

Example : a fair coin

x = outcome of coin toss

$x \sim Ber(p)$

$p \in [0,1]$

$H_0 : p = 0.5$

$H_1 : p \neq 0.5$

How to construct a test $D(x)$?

Suppose we have $X_1, X_2, \ldots, X_n \sim Ber(p)$

An intuitive test

$$D(x) = \begin{cases} \text{reject} & |x_n - 0.5| > c \\ \text{accept } H_0 & \text{o.w} \end{cases}$$

⇒ what c = ?

↗ type I error

If c too small, likely to reject H_0 when H_0 is true.

If c too large, likely to accept H_1 even if H_0 false.

↘ type II error

It's impossible to get rid of type-I & type-II errors

but we can calculate the probability of error.

Recall:

size = P(type-I error)

power = $1 - P$(type-II error)

Next Step

$n = 10$, $c = 0.15$

$$\delta(x) = \begin{cases} \text{reject} & \text{if } |\bar{x}_n - 0.5| > 0.15 \\ \text{accept} & \text{o.w} \end{cases}$$

$P(\text{Type I error})$ $p = 0.5$

$P(\delta(x) = \text{rejects})$

$\bar{x}_n \in \{0, 0.1, 0.2, \ldots, 1\}$

$P(\text{reject}) = P(|\bar{x}_n - 0.5| > 0.15)$

$= P(x_n \in \{0, 0.1, 0.2, 0.3, 0.7, 0.8, 0.9, 1\})$

$= 1 - P(x_n \in \{0.4, 0.5, 0.6\})$

$P(\text{type I error})$

$= 1 - P(\bar{x}_n = 0.4) - P(\bar{x}_n = 0.5) - P(\bar{x}_n = 0.6)$

$= 1 - C_4^{10} 0.5^4 0.5^6 - C_5^{10} 0.5^5 0.5^5 - C_6^{10} 0.5^6 0.5^4$

Remark:

$P(\text{type-I error})$ depends on c

e.g. $c = 0.08$ $(n = 10)$

$|\bar{x}_n - 0.5| > 0.08$ rejects

when $n=10$ $\{|\bar{x}_n - 0.5| < 0.08\} = \{\bar{x}_n = 0.5\}$

size $= 1 - P(\bar{x}_n = 0.5)$

c越小 → size越大 #

P(type Ⅱ error)

$p \neq 0.5$ $n=10$, $c=0.15$

P(D(x) Accepts)

$= P(\bar{x}_n = 0.6) + P(\bar{x}_n = 0.5) + P(\bar{x}_n = 0.4)$

$C_6^{10} p^6 (1-p)^4 + C_5^{10} p^5 (1-p)^5 + C_4^{10} p^4 (1-p)^6$

Remark:

1. P(type Ⅱ error) depends on c

⇒ c越大 P(type Ⅱ error)越大

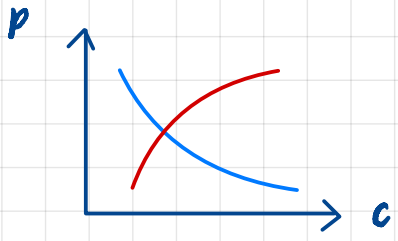
2. P(type Ⅱ error) depends on p

① $p = 0.49$

② $p = 0.01$

3. There is a trade-off between

P(type-Ⅰ) & P(type-Ⅱ)



Problem of exact calculations when $n = 1000$

$$C_{500}^{1000} \, 0.5^{500} \, 0.5^{500}$$

Such terms will appear . . .

Error probability approximation with CLT

CLT

$$\sqrt{n}\,(\bar{x}_n - p) \xrightarrow{d} N(0, p(1-p))$$

$$\Rightarrow \frac{\sqrt{n}\,(\bar{x}_n - p)}{\sqrt{p(1-p)}} \approx N(0,1) \quad \text{by CLT } (n \to \infty)$$

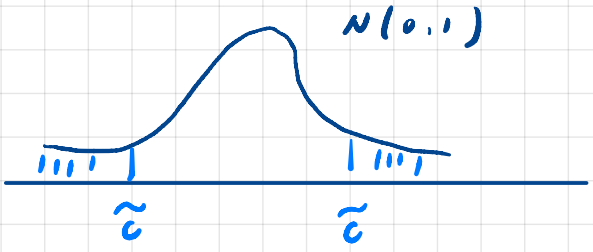
$$p(\text{type I error}) = p(b(x) = \text{reject})$$

$p = 0.5$

$$= p(|\bar{x}_n - 0.5| > c)$$

$$= p\left(\frac{\sqrt{n}\,|\bar{x}_n - 0.5|}{\sqrt{0.5(1-0.5)}} > \frac{\sqrt{n}\,c}{0.5}\right)$$

$$= p\left(|N(0,1)| > \frac{\sqrt{n}\,c}{0.5}\right) = 2\Phi\left(\frac{-\sqrt{n}\,c}{0.5}\right)$$



$$p(|N(0,1)| > \tilde{c}) = 2\,p(N(0,1) < -\tilde{c}) = 2\,\Phi(-\tilde{c})$$

Φ : $N(0,1)$ CDF

(when n large)

Given n, c

$P(\text{type I error})$

$= 2\Phi\left(\frac{-\sqrt{n}\,c}{0.5}\right)$ want $= 0.05$

when $n = 1000$ want $P(\text{type I}) = 0.05$

solve ~~$2\Phi(-2\sqrt{1000}\cdot c) = 0.05$~~

0.025

take $\Phi^{-1}(\dots)$

$$-2\sqrt{1000}\,c = \Phi^{-1}(-0.025)$$

$$c = \frac{\Phi^{-1}(-0.025)}{-2\sqrt{1000}}$$

$$= \frac{-1.96}{-2\sqrt{1000}} \approx \frac{-1.96}{-64} \doteq 0.03 = 3\%$$

case study：藍白合

Assume our goal is to find the best pair

HK v.s KH

侯柯 v.s 柯侯

Q1 HK v.s 賴

$X_{HK}=1$, $X_{HK} \sim Ber(p_{HK})$

p_{HK}：侯柯支持率

Q2 KH v.s 賴

$X_{KH} \sim Ber(p_{KH})$

$$\begin{cases} H_0: p_{HK} > p_{KH} \quad (=) \text{讓KMT除非強力證據說柯侯強} \\ H_1: p_{HK} < p_{KH} \end{cases}$$

Recall: we always prioritize $P(\text{type I error}) = 0.5$

侯柯強 but test says 柯侯強

$H_0: p_{HK} > p_{KH}$

A test if $\hat{p}_{HK} - \hat{p}_{KH} < c$, $c > 0$ we reject

$P(\text{type I error}) = P(\hat{p}_{HK} - p_{KH} < -c) \overset{\text{Want}}{=} 0.05$

朱立倫：If $\hat{p}_{HK} + 0.03 - (p_{KH} - 0.03) < 0$

$\Leftrightarrow p_{HK} - p_{KH} < -0.06$

$P(|\hat{p}_{HK} - p_{HK}| > 0.03) = 0.05$ } True

$P(|\hat{p}_{KH} - p_{KH}| > 0.03) = 0.05$

~~//~~

"So" $P(\hat{p}_{HK} - \hat{p}_{KH} < 0.06) = 0.05$

0.05 柯文哲 = 0.1（柯讓太少）

= 0.01（柯讓太多）

協議：

誤差內=侯柯+1

is ambiguous

1、什麼民調？

2、誤差幾%？

3、H_0

$H_0: p_{HK} > p_{KH}$

$H_0: p_{HK} > p_{L.KH}$

$p_{KH} > p_{L.KH}$

Error calculation with concentration inequality

Concentration inequality

"It's unlikely that X deviate from μ a lot

e.g. chebyshev inequality

$$P(|X-EX|>M) \le Var(X)/M$$

(上界)

X "concentrates" on EX

Hoeffding inequality for bernoulli r.v.

Thm. Let $X_1, X_2, \dots, X_n \sim Ber(p)$

$$P(|\bar{x}_n - p| > \varepsilon) \le 2e^{-2n\varepsilon^2} \quad \forall p$$

($\varepsilon = 0.1$)	$n=10$	$n=100$	$n=200$	$n=500$
$2e^{-2n\varepsilon^2}$	1.63	0.27	0.037	$\le 10^{-6}$

沒用

機率 < 1

n越大...?

$P(\text{type I error}) = P(|\bar{x}_n - 0.5| > c) \le 2e^{-2nc^2} \overset{\text{want}}{=} 0.05$

by Hoeffding

$$e^{-2nc^2} = 0.05/2$$

$$-2nc^2 = \ln(0.025)$$

$$c^2 = \ln(0.025)/-2n$$

when $n = 1000$, $c \approx 0.05$

n 夠大，樣本不會離真實值太遠

predict \ Actual	cat	dog
cat	TP	FP
dog	FN	TN

TP = true positive

$$\text{precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$