Hypothesis Testing

H_0 : statement 1 真相只有1個，是哪個？

H_1 : statement 2

We have data $X = x_1, x_2, \cdots, x_n$

a test is data-driven $D(X) \in \{ \text{Reject } H_0, \text{Accept } H_0 \}$

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

$X \sim \text{Bernoulli}(p)$, $p \in [0,1]$

$$\begin{cases} H_0 : p = 0.5 \\ H_1 : p \neq 0.5 \end{cases}$$

Suppose we have :

$X_1, X_2, \cdots, X_n \sim \text{Ber}(p)$

An intuitive test :

$$D(x) = \begin{cases} \text{reject } H_0, \text{ when } \left|\bar{X}_n - 0.5\right| > C \\ \text{Accept } H_0, \text{ 反之} \end{cases}$$

C 怎麼取？

⇒ 若 c 太小，很容易就會拒絕 H_0 even if H_0 true (type 1 error)

⇒ 若 c 太大，很難拒絕 H_0 even if H_0 false (type 2 error)

↓

無可避免

∵ 資料是 random 的

但可控制

Recall :

size = P(type 1 error) ⇒ H_0 為真，卻拒絕 H_0

power = 1 − P(type 2 error) ⇒ H_0 為假，拒絕 H_0

given threshold c ， 計算 size、power

當 $n=10$, $C=0.15$

$$D(x) = \begin{cases} \text{reject}, \text{ if } \left|\bar{X}_n - 0.5\right| > 0.15 \\ \text{accept}, \text{ 反之} \end{cases}$$

size = $P(\text{Type 1 error}) = P(D(x) = \text{rejects} \mid p = 0.5)$

$\bar{X}_n = \{0, 0.1, 0.2, \cdots, 1\}$

$P(\text{reject}) = P(\left|\bar{X}_n - 0.5\right| > 0.15)$, $\bar{X}_n \mid \text{reject} \in (0, 0.1, 0.2, 0.3, 0.7, 0.8, 0.9, 1)$

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

$\therefore P(\text{type 1 error}) = 1 - P(\bar{X}_n = 0.4) - P(\bar{X}_n = 0.5) - P(\bar{X}_n = 0.6)$

$= 1 - \binom{10}{4} 0.5^4 0.5^6 - \binom{10}{5} 0.5^5 0.5^5 - \binom{10}{6} 0.5^6 0.5^4$

$\Rightarrow$ Remark:

犯 type 1 error 的機率 取決於 C (C 越小 type error 越大)

e.g. $C = 0.08$, $n = 10$

rejects: $|\bar{X}_n - 0.5| > 0.08$, $\bar{X}_n | \text{reject} \in (0, 0.1, \cdots 0.4, 0.6 \cdots 1)$

$P(|\bar{X}_n - 0.5| > 0.15)$

$= 1 - P(x_n \in 0.5)$

$\therefore P(\text{type 1 error}) = 1 - P(\bar{X}_n = 0.5) = 1 - \binom{10}{5} 0.5^5 0.5^5$ 變大

P(type 2 error)

$\boxed{p \neq 0.5}$ $n=10, \ c=0.15$

$$P\left(D(X) \text{ accepts} \mid H_0 \text{ false} \right)$$
$$= P(\bar{X}_n = 0.6) + P(\bar{X}_n = 0.5) + P(\bar{X}_n = 0.4)$$
$$= \binom{10}{6} p^6 (1-p)^4 + \binom{10}{5} p^5 (1-p)^5 + \binom{10}{4} p^4 (1-p)^6$$

Remark:

P(type 2 error) 也取決於 c

$c \uparrow \Rightarrow P(\text{type 2 error}) \uparrow$

Remark:

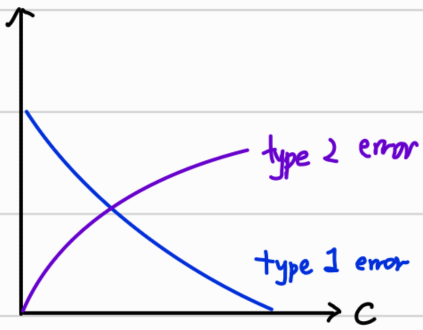
P(type 2 error) 也取決於 p

① $p = 0.49$

② $p = 0.01$

Remark:

$c \downarrow$ P(type 1 error) $\uparrow$
$c \uparrow$ P(type 2 error) $\uparrow$ $\Rightarrow$ tradeoff



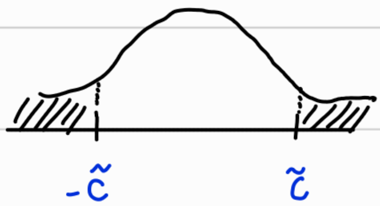
當 n 很大時，以上步驟無法用手算

Error probability approximation with CLT

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

$$\Rightarrow \frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1-p)}} \xrightarrow{d} N(0,1) \qquad (n \to \infty)$$

$$\begin{aligned} P(\text{type 1 error}) &= P(D(x) = \text{reject} \mid H_0 \text{ true}) \\ &= P\left(\frac{\sqrt{n}\,|\bar{X}_n - 0.5|}{\sqrt{0.5(1-0.5)}} > \frac{\sqrt{n}\, c}{0.5}\right) \\ &\xrightarrow{d} P\left(|N(0,1)| > \frac{\sqrt{n}\cdot c}{0.5} = \tilde{c}\right) \end{aligned}$$



$$= 2P(N(0,1) < -\tilde{c})$$

$$= 2\underbrace{\Phi}_{\text{standard normal CDF}}(-\tilde{c}) \quad \#$$

所以，當 n 很大時

given n, c

$p(\text{type 1 error}) = 2\Phi\left(\frac{-\sqrt{n}c}{p} \middle| H_0\text{為真}\right)$

e.g. when $n=1000$，我們要求 $p(\text{type 1 error}) = 0.05$

$$2\Phi\left(\frac{-\sqrt{1000}\cdot c}{0.5}\right) = 0.05$$

$$\Phi\left(-2\sqrt{1000}\,c\right) = 0.025$$

$$\Rightarrow c = \frac{\Phi^{-1}(0.025)}{-2\sqrt{1000}} = \frac{-1.96}{-2\sqrt{1000}} \approx \frac{1.96}{64} \fallingdotseq 0.03$$

Error calculation with concentration inequalities

Concentration inequality

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

e.g. **Chebshev Inequality**

$$P(|X-E(X)|>M) \le \frac{Var(X)}{M}$$

$\Rightarrow$ X "concentrates on" E(X)

Hoeffding inequality for Bernoulli RV

令 $X_1, X_2, \cdots, X_n \sim Bernoulli(p)$

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

e.g. $(\varepsilon = 0.1)$

	$n=10$	$n=100$	$n=200$	$n=500$
$2e^{-2n\varepsilon^2}$	~~1.63~~ 不合理	0.27	0.037	$\le 10^{-6}$

$$P(\text{type 1 error}) = \underbrace{P\left(\left|\bar{X}_n - 0.5\right| > c\right) \le 2e^{-2nc^2}}_{\text{Hoeffding}} = 0.05$$

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

$$\Rightarrow \quad -2nc^2 = \ln 0.025 = \ln 2.5 - 2$$

when $n=1000$, $c \cong 0.03$

(假設 cat 為陽性)

TP : true positive

		Actual: Cat	Actual: Dog
predict	Cat	TP	FP
predict	Dog	FN	TN

指標：

$$\text{precision} = \frac{TP}{TP + FP} \qquad \left(\frac{\text{actual 陽} \mid \text{predict 陽}}{\text{predict 陽性}} \right)$$

$$\text{recall} = \frac{TP}{TP + FN} \qquad \left(\frac{\text{predict 陽} \mid \text{actual 陽}}{\text{actual 陽性}} \right)$$

Case Study 藍白合：

assume our goal is to find the best pair

HK vs. KH

問卷：

Question ①：HK vs. 賴？

$X_{HK} = 1$, $X_{HK} \sim \text{Bernoulli}(P_{HK})$

HK的支持率

Question ②：KH vs. 賴？

$X_{KH} = 1$, $X_{KH} \sim \text{Bernoulli}(P_{KH})$

KH支持率

把HK放在H0 ⇒ 讓國民黨

H0為真，但拒絕H0

$$\begin{cases} H_0: & P_{HK} > P_{KH} \\ H_1: & P_{HK} < P_{KH} \end{cases}$$

除非有強力證據 KH強

Recall：we always prioritize $P(\text{type 1 error}) = 0.05$

（HK強，但test告訴我KH強）

if $\hat{P}_{HK} - \hat{P}_{KH} < c$, $c > 0$

⇒ we reject H_0

$$P(\text{type 1 error}) = P(\hat{P}_{HK} - \hat{P}_{KH} < -c) = 0.05$$

if $\hat{P}_{HK} + 0.03 - (P_{KH} - 0.03) < 0$

$\Leftrightarrow \quad P_{HK} - P_{KH} < -0.06$

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

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

by 朱立倫：

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

都是錯的

以上結論並不 imply 這兩人的 proposal

by 柯文哲：

"So", $P(\hat{P}_{HK} - \cancel{\hat{P}_{KH}} < -0.03) = 0.05$

規則寫的太模糊

① 哪家民調?

② 誤差 = ?

③ H_0 根本不清楚 $\Rightarrow$ 是 $H_0: P_{HK} > P_{KH}$

還是 $H_0: P_{HK} - P_{L\cdot HK} > P_{KH} - P_{L\cdot KH}$