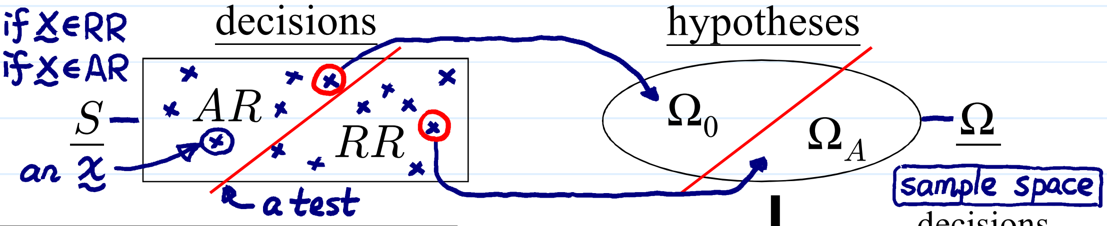
• Neyman-Pearson paradigm --- concept and procedure of hypothesis testing

Definition 7.2 (test, rejection region, acceptance region, test statistic, TBp. 331)

- 檢定 A **test** is a rule defined on sample space of data (denoted as S) that specifies:
  all possible outcomes
  1. for which sample values $\mathbf{x} \in S$, the decision is made to accept H_0 接受
  2. for which sample values $\mathbf{x} \in S$, reject H_0 in favor of H_A 棄卻
- The subset of the sample space for which H_0 will be rejected is called the **rejection region**, denoted as RR. 棄卻域
- The complement of the rejection region is called the **acceptance region**, denoted as $AR = S \setminus RR = RR^c$. 接受域

$$\phi(\underset{\sim}{x}) = \begin{cases} 1, & \text{if } \underset{\sim}{x} \in RR \\ 0, & \text{if } \underset{\sim}{x} \in AR \end{cases}$$



Note. Two types of errors may be incurred in applying this paradigm

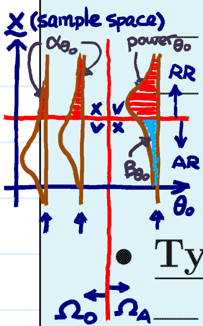
parameter space

		decisions: accept H_0	decisions: reject H_0
hypotheses	H_0 is true	Bingo!	Type I error
hypotheses	H_A is true	Type II error	Bingo!

Definition 7.3 (type I error, significance level, type II error, power, TBp. 331)

Let Θ_0 denote the true value of parameters. random

- **Type I error**: reject H_0 when H_0 is ture, i.e., $\underbrace{\mathbf{x} \in RR}_{\text{an event}}$ when $\Theta_0 \in \Omega_0$. not random
  - α_{Θ_0} = probability of Type I error = $P(RR|\Theta_0)$, where $\Theta_0 \in \Omega_0$.
  - **Significance level** α: the maximum (or the least upper bound) of the Type I error probability, i.e., $\alpha = \max_{\Theta \in \Omega_0} \alpha_\Theta$ or $\alpha = \sup_{\Theta \in \Omega_0} \alpha_\Theta$.



- **Type II error**: accept H_0 when H_A is ture, i.e., $\underbrace{\mathbf{x} \in AR}_{\text{an event}}$ when $\Theta_0 \in \Omega_A$
  - β_{Θ_0} = probability of Type II error = $P(AR|\Theta_0)$, where $\Theta_0 \in \Omega_A$
  - For $\Theta_0 \in \Omega_A$, **power**$_{\Theta_0} = 1 - \beta_{\Theta_0} = P(RR|\Theta_0)$ = probability of rejecting H_0 when H_A is ture

right decision when H_A true

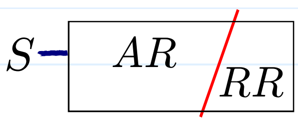
Question 7.4 (dilemma between type I and type II errors)

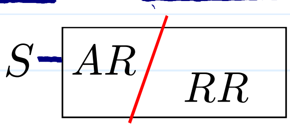
An ideal test would have $\alpha_\Theta = 0$, for $\Theta \in \Omega_0$ and $\beta_\Theta = 0$, for $\Theta \in \Omega_A$; but this can be achieved only in trivial cases. (Example?) $\begin{cases} \Omega_0 = \{\text{uniform}(0,1)\} \\ \Omega_A = \{\text{uniform}(2,3)\} \end{cases}$

In general, when α decreases, β increases, and vice versa.

Q: Can we find test s.t. α is minimized, and β is minimized?

smaller α, larger β





larger α, smaller β

A solution to the dilemma

Strong evidence: H_A is true → reject H_0. Otherwise → accept H_0.

The Neyman-Pearson approach imposes an asymmetry between H_0 and H_A: answer to question 7.3 in LNp.2

1. the significance level α (Type I error) is fixed in advance, usually at a rather small number (e.g. 0.1, 0.05, and 0.01 are commonly used), and → protect H_0
2. then try to construct a test yielding a small value of β_Θ (Type II error) for $\Theta \in \Omega_A$.

$\underset{\text{tests}}{\arg\min}\ \beta_\theta$ subject to tests with $\sup_{\theta\in\Omega_0} \alpha_\theta \le \alpha$ (level-α tests)

Example 7.2 (cont. Ex. 7.1 item 1 (LNp.3), TBp.330-331)

- Testing the value of the parameter p of a Bernoulli distribution with 10 trials. Let X (Data) be the number of successes, then $X \sim B(10, p)$ ← statistical modeling

$\Omega = \{B(10, 0.5), B(10, 0.7)\}$

$$H_0 : p = 0.5 \quad \text{vs.} \quad H_A : p = 0.7$$

$T(X_1, \cdots, X_{10})$

H_0 0 1 2 3 4 5 6 7 8 9 10 H_A

more support on H_A

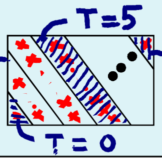
more extreme ← How to quantify it?

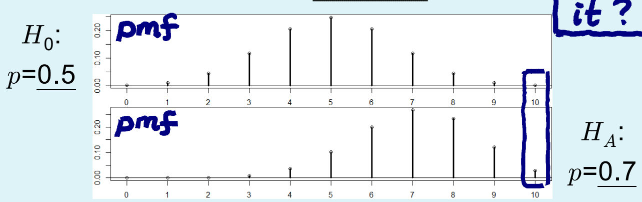
- *Different observations* of X give *different support* on H_A

$X_1, \cdots, X_{10} \sim B(1, p)$ → $X = T(X_1, \ldots, X_{10})$

$T = X = \sum_{i=1}^{10} X_i$

$S = \{(x_1, \cdots, x_{10}) \mid x_i \in \{0,1\}, i = 1, \cdots, 10\}$





– In this case, larger observations of X give more support on H_A.

– Rejection region should consist of large values of X.

– The observations that give more support on H_A are called "*more extreme*" observations

most extreme : $X = 10$, 2nd most extreme : $X = 9$, …

- If the rejection region is $\{7, 8, 9, 10\}$, then – a test is formed

$X \sim B(10, 0.5)$ under H_0

$$\alpha = P(X \in \{7, 8, 9, 10\} \mid p = 0.5) = 1 - P(X \le 6 \mid p = 0.5) = 0.18.$$

RR, H_0, AR, H_0

- *Set the significance level as* 0.18.

– The probability of Type II error (i.e., β) is

$X \sim B(10, 0.7)$ under H_A

$$P(X \notin \{7, 8, 9, 10\} \mid p = 0.7) = P(X \le 6 \mid p = 0.7) = 0.35,$$

AR, H_A, AR, H_A

– and the power (i.e., $1 - \beta$) is

$$P(X \in \{7, 8, 9, 10\} \mid p = 0.7) = 1 - P(X \le 6 \mid p = 0.7) = 0.65.$$

RR, H_A, AR, H_A

- Suppose that we are interested in testing

$\Omega = \{p \mid p \in [0, 1]\}$ ← ($\le$)

$$H_0 : p = 0.5 \quad \text{vs.} \quad H_A : p > 0.5$$ ← $\Omega = \{p \mid p \in [0.5, 1]\}$

and the rejection region is still set to be $\{7, 8, 9, 10\}$:

– Note that β_p is a function of p: $X \sim B(10, p)$

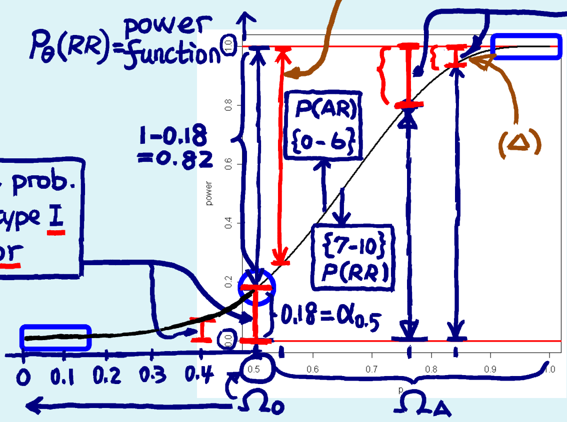
defined for $\begin{cases} p \in \Omega_A, \text{ and} \to 1-\beta_p \\ p \in \Omega_0 \to \alpha_p \end{cases}$

$$\beta_p = P(\underbrace{X \notin \{7, 8, 9, 10\}}_{AR} \mid p).$$

– power function $\equiv 1-\beta_p$. $= P(RR \mid p)$

– $1-\beta_p \to 1$ as $p \to 1$, and $1-\beta_p \to \alpha$ as $p \to 0.5$.

– $1-\beta_p \to 0$ as $p \to 0$.



β_p = prob. of type II error

Significance level $\sup_{p \le 0.5} \alpha_p = \alpha_{0.5} = 0.18$

α_p = prob. of type I error

Check (*) in the graph

P_θ(type II error) < ? for $\theta \in \Omega_A$

cf. P_θ(type I error) $\le \alpha$, $\theta \in \Omega_0$

Question 7.5

1. Which hypothesis get more *protection*? (H_0) What kind of protection?
2. Can the protection make P(Type I error) always smaller than P(Type II error), i.e., $\alpha < \beta_\Theta$ for any $\Theta \in \Omega_A$? No. check (Δ) in the graph