STAT201B Lecture 9 Hypothesis Testing

Logic

关于 Hypothesis testing 的定义的更详细的论述, 见 [STA3020 Lecture 14]

1 Hypothesis Testing 的相关定义

1.1 Definition: Hypothesis

Statistical Hypothesis 是一个关于 parameter (或 nonparametric models 中的 statistical functional) 的 statement

1.2 Definition: Hypothesis testing

Hypothesis test 满足:

- 将 parameter space Θ 分为两个 **disjoint sets** Θ_0 和 Θ_1
- 明确了一个 decision rule, 用于选择

$$H_0 : \theta \in \Theta_0 \quad and \quad H_1 : \theta \in \Theta_1$$

其中,

- H_0 被称为 null hypothesis
- H_1 被称为 alternative hypothesis

两个潜在的选择分别是

- Reject H_0
- Fail to reject H_0

1.3 Definition: Rejection region

Rejection region 满足:

- Decision (是否 reject H_0) 的选取取决于 sample $X = (X_1, \ldots, X_n)$ 是否落在 predefined rejection region R 中
- 通常有以下形式

$$R = \{x_1, \ldots, x_n : T(x_1, \ldots, x_n) > c\}$$

其中,

- T 被称为 **test statistics**
- c 被称为 **critical value**

Remark

- 当 H_0 为 True 的时候, 我们希望 data 落入 R 的概率尽可能小
- c 的取值通常取决于 n, Θ_0 与 Θ_1, α 的选取

1.4 Definition: Critical function (补充)

对于一个 rejection region 为 的 hypothesis testing, **critical function** 被定义为以下形式:

$$\phi_R(X) = \begin{cases} 1 & \text{if } X \in R \\ \gamma \in (0,1) & \text{if } X \in \partial R \\ 0 & \text{if } X \notin R \end{cases}$$

- 当 $\phi_R(X) = 1$ 时, 我们 reject H_0
- 当 $\phi_R(X) = \gamma$ 时, 我们 reject H_0 with probability γ

⚠ **Remark**

- 现实中的 critical function 常为 $\phi : \mathcal{X} \to \{0,1\}$, 定义 γ 是为了让 power 更大
- 现实中, 我们常常是先选择一个 critical function, 再由此 induce 出一个 rejection region:

$$\phi(x) \in [0,1],\ \forall x \in \mathcal{X} \quad \Rightarrow \quad R = \{x | \phi(x) = 1\},\ \partial R = \{x | 0 < \phi(x) < 1\}$$

1.5 Definition: Power function

对于一个 rejection region 为 R 的 hypothesis testing, **power function** 被定义为以下关于 θ 的函数:

$$\begin{aligned}
\beta(\theta) &= \mathbb{P}_\theta(\text{Reject } H_0) \\
&= \mathbb{P}_\theta(X \in R) \\
&= \begin{cases} \text{Type I error} & \text{if } \theta \in \Theta_0 \\ 1 - \text{Type II error} & \text{if } \theta \in \Theta_1 \end{cases} \\
&= \mathbb{E}_\theta[\phi(X)]
\end{aligned}$$

⚠ **Remark**

一个好的 hypothesis testing 应该满足:

- small $\beta(\theta)$ (尽可能接近 0) when $\theta \in \Theta_0$
- large $\beta(\theta)$ (尽可能接近 1) when $\theta \in \Theta_1$

Example

问题:

令 $X_1, \dots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$, 其中 σ^2 已知, 考虑 testing $H_0 : \mu = 0 \text{ versus } H_1 : \mu \neq 0$, 使用 rejection region

$$R = \{x_1, \dots, x_n : |\bar{X}_n| > c\}$$

求出 $\beta(\mu)$

解答:

$$\begin{aligned}
\beta(\mu) &= \mathbb{P}_\mu(|\bar{X}_n| > c) \\
&= \mathbb{P}_\mu(\bar{X}_n > c, \bar{X}_n < -c) \\
&= \mathbb{P}_\mu\left(\sqrt{n}\left(\frac{\bar{X} - \mu}{\sigma}\right) > \sqrt{n}\left(\frac{c - \mu}{\sigma}\right), \sqrt{n}\left(\frac{\bar{X} - \mu}{\sigma}\right) < \sqrt{n}\left(\frac{-c - \mu}{\sigma}\right)\right) \\
&= 1 - \Phi\left(\sqrt{n}\left(\frac{c - \mu}{\sigma}\right)\right) + \Phi\left(-\sqrt{n}\left(\frac{c - \mu}{\sigma}\right)\right)
\end{aligned}$$

当 $\mu = 0$ 时, Type I error $\beta(0)$ 为

$$\beta(0) = 1 - \Phi\left(\frac{\sqrt{n}c}{\sigma}\right) + \Phi\left(\frac{-\sqrt{n}c}{\sigma}\right) = 2\Phi\left(\frac{-\sqrt{n}c}{\sigma}\right)$$

Example

问题:

令 $X \sim Bin(5, p)$, 考虑 testing $H_0 : p \leq 1/2$ versus $H_1 : p > 1/2$, 我们使用以下两个 rejection regions:

$$R_1 = \{x : x = 5\} \qquad R_2 = \{x : x \geq 3\}$$

比较其对应的 power functions $\beta_1(p)$ 和 $\beta_2(p)$

解答:

其对应的 power function 为

$$\beta_1(p) = \mathbb{P}_p(x=5) = p^5$$
$$\beta_2(p) = \mathbb{P}_p(x \geq 3) = \sum_{i=3}^{5} \binom{5}{i} p^i (1-p)^{5-i}$$

现实中, 我们通常使用 R_1, 因为 $\beta_1(1/2)$ 相较于 $\beta_2(1/2)$ 更小, 更难 reject (Type I error 更小)

1.6 Definition: Type I error 和 Type II error

Logic

构造一个好的 test 有两种途径 (控制 Type I error 和控制 Type II error), 为了使 tests 之间的比较更加 better defined, 我们需要对 tests 作出一些限制 (控制 Type I error 相同, 比较 Type II error 的大小), 我们希望 find a test within a class that has large $\beta(\theta)$ for $\theta \in \Theta_1$

	Fail to Reject H_0	Reject H_0
H_0 true	Correct	**Type I error**
H_1 true	**Type II error**	Correct

- 对于 $\forall \theta_0 \in \Theta_0$, **Type I error** 被定义为:

$$\text{Type I error} = \mathbb{P}(\text{Reject } H_0 | \theta_0 \in \Theta_0) = \mathbb{P}_{\theta_0}(X \in R) = \mathbb{E}_{\theta_0}[\phi(X)]$$

- 对于 $\forall \theta_1 \in \Theta_1$, **Type II error** 被定义为:

$$\text{Type II error} = \mathbb{P}(\text{Fail to Reject } H_1 | \theta_1 \in \Theta_1) = 1 - \mathbb{P}_{\theta_1}(X \in R) = 1 - \mathbb{E}_{\theta_1}[\phi(X)]$$

1.7 Definition: Level α 和 Size α test

一个 test 的 **size** 被定义为:

$$\alpha = \sup_{\theta \in \Theta_0} \beta(\theta)$$

- 一个 test 被称为 **have level** α, 若它的 size $\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$
- 一个 test 被称为 **have size** α, 若它的 size $\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$

Remark

- 一个 test 的 size 为: H_0 为 true 时, test 拒绝 H_0 的最大的概率
- size α test 属于 level α test

Example

问题:
令 $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$, 其中 σ^2 已知, 考虑 testing $H_0 : \mu = 0 \text{ versus } H_1 : \mu \neq 0$, 使用 rejection region

$$R = \{x_1, \ldots, x_n : |\bar{X}_n| > c\}$$

求出使 test 为 size α test 的 c

解答:
在之前的例子中, 我们已经得到 Type I error $\beta(0)$ 为

$$\beta(0) = 1 - \Phi\left(\frac{\sqrt{n}c}{\sigma}\right) + \Phi\left(\frac{-\sqrt{n}c}{\sigma}\right) = 2\Phi\left(\frac{-\sqrt{n}c}{\sigma}\right)$$

令其等于 α, 可以得到:

$$2\Phi\left(\frac{-\sqrt{n}c}{\sigma}\right) = \alpha \quad \Rightarrow \quad \frac{\sqrt{n}c}{\sigma} = z_{1-\alpha/2} \quad \Rightarrow \quad c = \frac{(z_{1-\alpha/2})\sigma}{\sqrt{n}}$$

1.8 Point-null tests 和 confidence sets 的关系

Logic

关于 hypothesis testing 和 confidence sets 之间的 duality 的详细论述, 见 [STA3020 Lecture 21]

若我们希望检验 $H_0 : \theta = \theta_0$, 假设我们已经有一个关于 θ 的 $1 - \alpha$ confidence interval, 则一个 level α test 可以被构造为: reject H_0 若 θ_0 位于 interval 之外