STAT201B Lecture 11 p-value

1 P-value

Logic

关于 P-value 的详细论述, 见 STA3020 Lecture 14

1.1 Definition: p-value

若对于任意 $\alpha \in (0,1)$, 我们均有 size α test, 且 rejection region 为 $R_\alpha = \{x : T(x) \geq c_\alpha\}$,
则 p-value 被定义为:

$$\text{p-value} = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(T(X) \geq T(x))$$

其中 x 为 observed data, X 为 random variable 且 $X \overset{d}{=} x$

Remark

根据定义, p-value 表示 H_0 下, test statistic $T(X)$ 的值与 observed value 相等或更极端的最大概率

1.2 Definition: p-value 的等价定义

若对于任意 $\alpha \in (0,1)$, 我们均有 size α test, 且 rejection region 为 R_α,
则 p-value 被定义为:

$$\text{p-value} = \sup_{\theta \in \Theta_0} \inf\{\alpha : 0 \leq \alpha \leq 1, x \in R_\alpha\}$$

Remark

根据定义, p-value 表示在给定观测值的前提下, 使我们能 reject H_0 的最小 level

Example

1. 在 Wald test 的情形下, (approximated) p-value 为:

$$\text{p-value} = P_{\theta_0}(|W| > |w|) \approx P(|Z| > |w|) = 2\Phi(-|w|)$$

   其中 w 为统计量的观测值, 且 $Z \sim \mathcal{N}(0,1)$
2. 在 LRT with point null hypothesis 的情形下, 且其 limiting distribution 为 χ^2_{r-q} 时, (approximated) p-value 为:

$$\text{p-value} = P_{\theta_0}(\lambda(X) > \lambda(x)) \approx P(\chi^2_{r-q} > \lambda(x))$$

1.3 Theorem: Simple Null Hypothesis 下 p-value 的分布

若:

1. hypothesis testing problem 有 simple null hypothesis

$$H_0 : \theta = \theta_0 \quad v.s. \quad H_1 : \theta \in \Theta_1$$

2. rejection region 为 $R_\alpha = \{T(X) > c(\alpha)\}$

则 under $H_0 : \theta = \theta_0$, 有

$$\text{p-value} = 1 - F_{H_0}(T) \sim Unif(0,1)$$

Remark

因此, 若我们选择在 p-value 小于 α 的时候拒绝 H_0, 则犯 Type I error (在 H_0 为真的情况下拒绝 H_0) 的概率为 α

Proof

令 $\text{p-value} = 1 - F_{H_0}(T(X)) := Z$, 则有

$$\begin{aligned}
\mathbb{P}_{H_0}(Z \le z) &= \mathbb{P}_{H_0}(1 - F_{H_0}(T(X)) \le z) \\
&= \mathbb{P}_{H_0}(T(X) \ge F_{H_0}^{-1}(1-z)) \\
&= 1 - \mathbb{P}_{H_0}(T(X) \le F_{H_0}^{-1}(1-z)) \\
&= 1 - F_{H_0}(F_{H_0}^{-1}(1-z)) \\
&= z
\end{aligned}$$

因此,

$$\text{p-value} = Z \sim Unif(0,1)$$

2 Neyman-Pearson Theorem

Logic

关于 Simple Hypothesis 的 UMP 和 Neyman-Pearson Lemma 的详细论述, 见 STA3020 Lecture 15

关于 UMP 的详细论述, 见 STA3020 Lecture 15, STA3020 Lecture 16, STA3020 Lecture 17 , 内容包括:

- Uniformly most power test 的定义
- Simple hypothesis 的 UMP (Neyman-Pearson Lemma)
- Composite one-sided null hypothesis 的 UMP (Karlin-Rubin theorem)
- Composite two-sided null hypothesis 的 UMP (适用于 exponential family)
- Composite null (general) hypothesis 的 UMP

令 hypothesis 为:

$$H_0 : \theta = \theta_0 \quad v.s. \quad H_1 : \theta = \theta_1$$

若:

1. Test statistic T 为

$$T(X) = \frac{\mathcal{L}_n(\theta_1)}{\mathcal{L}_n(\theta_0)} = \frac{f(x_1, \dots, x_n; \theta_1)}{f(x_1, \dots, x_n; \theta_0)}$$

2. 在 $T > c$ 时 reject H_0
3. c 满足 $P_{\theta_0}(T > c) = \alpha$, 即确保 test 为 size- α test

则该 test 为 the most powerful, size α test, 即在所有 size α tests 中, 该 test 的 power $\beta(\theta_1)$ 最大

Remark: STA3020 中的 Neyman-Pearson Lemma

令:

1. test hypothesis 为 $H_0 : \theta = \theta_0 \quad v.s. \quad H_1 : \theta = \theta_1$
2. x 关于 θ_0, θ_1 的分布为: $f(x|\theta_i),\ i = \{0, 1\}$

若一个 test 满足以下条件:

1. critical function 的形式为:

$$\phi = \mathbf{1}\left(\frac{f(x|\theta_1)}{f(x|\theta_0)} > k\right) + \gamma \cdot \mathbf{1}\left(\frac{f(x|\theta_1)}{f(x|\theta_0)} = k\right) = \begin{cases} 1, & \text{if } \frac{f(x|\theta_1)}{f(x|\theta_0)} > k \\ \gamma, & \text{if } \frac{f(x|\theta_1)}{f(x|\theta_0)} = k \\ 0, & \text{if } \frac{f(x|\theta_1)}{f(x|\theta_0)} < k \end{cases}$$

2. 该 test 为 size alpha test:

$$\mathbb{E}_{\theta_0}[\phi(X)] = \alpha$$

则:

1. **Sufficiency**: 任意满足以上条件的 test 均为 UMP size α test
2. **Necessity**: 任意 UMP test 均满足以上条件 (except on a set $\mathcal{A}$ satisfying $P_{\theta_0}(X \in \mathcal{A}) = P_{\theta_1}(X \in \mathcal{A}) = 0$)

Example: Exponential family 的 UMP

例 1: 令 random sample $X = \{X_1, \cdots, X_n\}$ 的分布属于 one-parameter exponential family:

$$f(x|\theta) = c(\theta) \cdot h(x) \cdot \exp\{\theta \cdot T(x)\}$$

① Construct a UMP size α test for the hypothesis:

$$H_0: \theta = \theta_0 \quad v.s. \quad H_1: \theta = \theta_1 \quad (\theta_1 > \theta_0)$$

UMP test 的 test statistic 为:

$$R = \frac{\prod_{i=1}^n f(x_i|\theta_1)}{\prod_{i=1}^n f(x_i|\theta_0)} = \left(\frac{c(\theta_1)}{c(\theta_0)}\right)^n \exp\left\{(\theta_1 - \theta_0) \cdot \sum_{i=1}^n T(x_i)\right\}$$

$$\Rightarrow \phi = 1(R > k) + \gamma \cdot 1(R = k)$$

$$= 1\left\{\sum_{i=1}^n T(x_i) > k'\right\} + \gamma \cdot 1\left\{\sum_{i=1}^n T(x_i) = k'\right\} \quad (\textstyle\sum_{i=1}^n T(x_i) \text{ 的分布可能离散})$$

其中 k' is determined by

$$E_{\theta_0}\phi(X) = P\left\{\sum_{i=1}^n T(x_i) > k'\right\} + \gamma \cdot P\left\{\sum_{i=1}^n T(x_i) = k'\right\} = \alpha$$

② Construct a UMP size α test for the hypothesis:

$$H_0: \theta = \theta_0 \quad v.s. \quad H_1: \theta > \theta_0$$

(此处不能直接使用 Neyman-Pearson Lemma，但可以利用上一问的结果)

先 look at $H_0: \theta = \theta_0 \quad v.s. \quad H_1: \theta = \theta_2, \ \forall \theta_2 > \theta_0$

根据上一问的结果，有 UMP size α test:

$$\phi = 1\left\{\sum_{i=1}^n T(x_i) > k'\right\} + \gamma \cdot 1\left\{\sum_{i=1}^n T(x_i) = k'\right\}$$

γ, k' satisfying $P\left\{\sum_{i=1}^n T(x_i) > k'\right\} + \gamma \cdot P\left\{\sum_{i=1}^n T(x_i) = k'\right\} = \alpha$

因此,

对 $\forall \theta_2 > \theta_0$ ($\theta_2 \in \Theta_1$), ϕ 均为 UMP, 且 ϕ independent with θ_2 ($\sum_{i=1}^n T(x_i) \perp \theta_2$)

$\Rightarrow$ 对 $\forall \theta_2 > \theta_0$ ($\theta_2 \in \Theta_1$), $\beta_\phi(\theta_2) \geq \beta_{\phi'}(\theta_2)$, $\forall \phi'$ being size α

$\Rightarrow$ ϕ 为 UMP for $H_0: \theta = \theta_0 \quad v.s. \quad H_1: \theta > \theta_0$ (根据 UMP 的定义)

Example: 双侧检验不存在 UMP

例2：（Two side hypothesis 不存在 UMP test）

考虑 random sample $X=\{X_1,\cdots,X_n\}\sim N(\theta,1)$，证明以下 two sided hypothesis 没有 UMP：

$H_0:\theta=\theta_0$ v.s. $H_1:\theta\neq\theta_0$

① 首先研究 $H_0:\theta=\theta_0$ v.s. $H_1^{(1)}:\theta=\theta_1$ （Assume $\theta_1>\theta_0$）

UMP test 的 test statistic 为：

$$R=\frac{\prod_{i=1}^n f(x_i|\theta_1)}{\prod_{i=1}^n f(x_i|\theta_0)}=\frac{(2\pi)^{-\frac{n}{2}}\exp(-\frac{1}{2}\sum_{i=1}^n(x_i-\theta_1)^2)}{(2\pi)^{-\frac{n}{2}}\exp(-\frac{1}{2}\sum_{i=1}^n(x_i-\theta_0)^2)}=\exp\{-\frac{n}{2}(\theta_1^2-\theta_0^2)\}\exp\{\underbrace{(\theta_1-\theta_0)}_{>0}\sum_{i=1}^n x_i\}$$

$$\Rightarrow \phi=1(R>k)+\gamma\cdot 1(R=k)$$

$$=1\{\bar{X}>k'\}$$ （$T(X):=\bar{X}$ 服从连续分布，故可令 $\gamma=0$）

其中 k' is determined by

$$E_{\theta_0}\phi(X)=P\{\bar{X}>k'\}=\alpha$$

$\Rightarrow k'=\theta_0+\frac{1}{\sqrt{n}}\cdot z_\alpha$ 其中 z_α 为 $N(0,1)$ 的 α upper quantile （Under H_0, $\bar{X}\sim N(\theta,\frac{1}{n})$）

因此 $H_0:\theta=\theta_0$ v.s. $H_1':\theta>\theta_0$ 的 UMP 为

$$\phi=1\{\bar{X}>\theta_0+\frac{1}{\sqrt{n}}\cdot z_\alpha\}$$ （因为 $k'\perp\theta_1$，因此可从 $H_1^{(1)}$ 推广到 H_1'）

② 接着研究 $H_0:\theta=\theta_0$ v.s. $H_1^{(2)}:\theta=\theta_2$ （Assume $\theta_2<\theta_0$）

UMP test 的 test statistic 为：

$$R=\frac{\prod_{i=1}^n f(x_i|\theta_2)}{\prod_{i=1}^n f(x_i|\theta_0)}=\frac{(2\pi)^{-\frac{n}{2}}\exp(-\frac{1}{2}\sum_{i=1}^n(x_i-\theta_2)^2)}{(2\pi)^{-\frac{n}{2}}\exp(-\frac{1}{2}\sum_{i=1}^n(x_i-\theta_0)^2)}=\exp\{-\frac{n}{2}(\theta_2^2-\theta_0^2)\}\exp\{\underbrace{(\theta_2-\theta_0)}_{<0}\sum_{i=1}^n x_i\}$$

$$\Rightarrow \phi'=1(R>k)+\gamma\cdot 1(R=k)$$

$$=1\{\bar{X}<k'\}$$ （$T(X):=\bar{X}$ 服从连续分布，故可令 $\gamma=0$）

其中 k' is determined by

$$E_{\theta_0}\phi'(X)=P\{\bar{X}<k'\}=\alpha$$

$\Rightarrow k'=\theta_0+\frac{1}{\sqrt{n}}\cdot z_{1-\alpha}$ 其中 $z_{1-\alpha}$ 为 $N(0,1)$ 的 $1-\alpha$ upper quantile （Under H_0, $\bar{X}\sim N(\theta,\frac{1}{n})$）

因此 $H_0:\theta=\theta_0$ v.s. $H_1'':\theta<\theta_0$ 的 UMP 为

$$\phi'=1\{\bar{X}<\theta_0+\frac{1}{\sqrt{n}}\cdot z_{1-\alpha}\}$$ （因为 $k'\perp\theta_1$，因此可从 $H_1^{(2)}$ 推广到 H_1''）

（很显然 $\phi\neq\phi'$，即 $E_{\theta_0}|\phi-\phi'|>0$，接下来我们仅需利用 UMP 的定义和 NP theorem 的 necessity 即可反证出 UMP 不存在）

③ Argue by contradiction：假设 $H_0:\theta=\theta_0$ v.s. $H_1:\theta\neq\theta_0$ 的 UMP 存在，且为 ϕ^*

根据 UMP 的定义（$\theta\neq\theta_0$ 时 ϕ^* 的 type-II error 最小），

$$\phi^* \text{也是}\begin{cases}H_0:\theta=\theta_0 \text{ v.s. } H_1':\theta>\theta_0 \text{ 的 UMP}\\ H_0:\theta=\theta_0 \text{ v.s. } H_1'':\theta<\theta_0 \text{ 的 UMP}\end{cases}$$

根据 NP theorem 的 necessity，有

$$\begin{cases}\phi^*=\phi \text{ a.s.} \iff E_{\theta_0}|\phi^*-\phi|=0\\ \phi^*=\phi' \text{ a.s.} \iff E_{\theta_0}|\phi^*-\phi'|=0\end{cases}$$