STAT201B Lecture 10 Wald Test and LRT

1 Wald Test

> **Logic**
>
> 关于 MLE 的 Wald test 的详细描述, 见 [STA2004 Lecture 12]

1.1 Wald test 的情景设置

1. **point-null hypothesis:** 检验

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

2. **asymptotic normal estimator:** Under null hypothesis $\theta=\theta_0$, θ 的 estimator $\hat{\theta}_n$ 服从 asymptotic normal distribution:

$$\frac{\hat{\theta}_n-\theta_0}{\hat{se}(\hat{\theta}_n)}\xrightarrow{d}\mathcal{N}(0,1)$$

> **Remark**
>
> Asymptotic normality 通常来自:
>
> 1. MLE 的 asymptotic normality
> 2. CLT

1.2 Wald test

- Wald test 的 test statistic 为:

$$T=\left|\frac{\hat{\theta}_n-\theta_0}{\hat{se}(\hat{\theta}_n)}\right|$$

- size α Wald test 的 rejection region 为:

$$T>z_{\alpha/2}$$

- Wald test 的 $1-\alpha$ confidence interval 为:

$$[\hat{\theta}_n-z_{\alpha/2}\cdot\hat{se}(\hat{\theta}_n),\ \hat{\theta}_n+z_{\alpha/2}\cdot\hat{se}(\hat{\theta}_n)]$$

1.3 Generalized Wald test (multi-parameter cases)

令:

1. $\hat{\theta}_n$ 为 θ 的 MLE
2. g 为一个 invertible function

则 Wald test 的构建可以基于:

$$\frac{g(\hat{\theta})-g(\theta_0)}{\hat{se}(g(\hat{\theta}_n))}\xrightarrow{d}\mathcal{N}(0,1)$$

其中 $\hat{se}(g(\hat{\theta}_n))$ 可以通过 Delta method 得到

1.4 Generalized Wald test (non-parametric cases)

令 $\theta=T(F)$, 其中 F 为 unknown distribution

1. 若 T **为 linear functional**, 则 plug-in estimator 为 a mean of iid random variables:

$$T(\hat{F}_n) = \int r(x) d\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} r(X_i)$$

此时可以使用 CLT 得到 asymptotic normality

2. 若 T **为 non-linear functional**, 则可以使用 bootstrap 来近似 $\hat{se}(\hat{\theta}_n)$, 若可以证明 asymptotic normality, 则可以使用 Wald test

Remark

现实中, 对于 non-linear functional, 即便 asymptotic normality 无法证明, 仍然可以使用 Wald test, 但需要承担风险

Example

问题:
若 $X \sim Bin(m, p_1)$, $Y \sim Bin(n, p_2)$, 构建一个 size α Wald test for $H_0 : p_1 = p_2$

解答:

令 $t = p_1 - p_2$, 则 null hypothesis 转化为: $H_0 : t = 0$,
由于 p_1, p_2 的 MLE 为

$$\hat{p}_1 = \frac{X}{m}, \quad \hat{p}_2 = \frac{Y}{n}$$

根据 MLE 的 equivariance property, 有

$$\hat{t} = \hat{p}_1 - \hat{p}_2 = \frac{X}{m} - \frac{Y}{n}$$

注意到:

$$Var[\hat{t}] = p_1(1-p_1) + p_2(1-p_2)$$

因此,

$$T = \left|\frac{\hat{t}}{\hat{se}(\hat{t})}\right| = \left|\frac{\hat{p}_1 - \hat{p}_2}{\hat{p}_1(1-\hat{p}_1) + \hat{p}_2(1-\hat{p}_2)}\right|$$

Reject null hypothesis 若

$$T > z_{\alpha/2}$$

Example

问题:
令:

1. $F(u, v)$ 为两个 random variables U 和 V 的 joint distribution
2. $\theta = T(F) = \rho(U, V) = \frac{\mathbb{E}[(U-\mathbb{E}[U])(V-\mathbb{E}[V])]}{\sqrt{Var[U]Var[V]}}$, 其中 ρ 表示 correlation

求 a size α Wald test for $H_0 : \rho = 0$ using the plug-in estimator and the bootstrap

解答:

plug-in estimator 为:

$$\hat{\rho} = \frac{\frac{1}{n}\sum_{i=1}^{n}(U_i - \bar{U}_n)(V_i - \bar{V}_n)}{\hat{se}(U)\hat{se}(V)}$$

2 Likelihood Ratio Test (LRT)

Logic

关于 LRT 的详细描述, 见 STA3020 Lecture 18

2.1 Likelihood ratio test (LRT) 的 test statistic

令:

1. random sample $X = \{X_1, \dots, X_n\} \overset{i.i.d.}{\sim} f(x|\theta)$
2. hypothesis 为:

$$H_0 : \theta \in \Theta_0 \quad v.s. \quad H_1 : \theta \in \Theta_1$$

   其中 $\Theta_0 \cap \Theta_1 = \emptyset$, $\Theta_0 \cup \Theta_1 = \Theta$

则 LRT 的 test statistic 为:

$$T(X) = \frac{\sup_{\theta \in \Theta} \mathcal{L}_n(\theta)}{\sup_{\theta \in \Theta_0} \mathcal{L}_n(\theta)}$$

若 $\hat{\theta}_n$ 为 MLE, $\hat{\theta}_{n,0}$ 为 MLE restricting $\theta \in \Theta_0$, 则

$$T(X) = \frac{\mathcal{L}_n(\hat{\theta}_n)}{\mathcal{L}_n(\hat{\theta}_{n,0})}$$

Remark

1. LRT 的 test statistic 满足: $T(X) \geq 1$
2. LRT 的 rejection region 形如:

$$R = \{x : T(x) > c\}$$

   若 $T(X)$ 的值很大, 则表示在 Θ_1 中, 有某个 θ 的 likelihood 比 Θ_0 中的 likelihood 都要大很多

Example

问题:
令:

1. $X_1, \dots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\theta, 1)$
2. $H_0 : \theta = \theta_0 \quad v.s. \quad H_1 : \theta \neq \theta_0$

求:

1. $T(X)$
2. simplified expression for the form of rejection region
3. $\theta = \theta_1$ 时的 power
4. size α LRT

解答:
显然, θ 的 MLE 为 $\bar{X}_n$, 因此 $T(X)$ 为

$$\begin{aligned}
T(X) &= \frac{\left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot \exp\left\{-\frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{2}\right\}}{\left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot \exp\left\{-\frac{\sum_{i=1}^n (X_i - \theta_0)^2}{2}\right\}} \\
&= \exp\left\{-\frac{\sum_{i=1}^n X_i^2 - n\bar{X}_n^2 - \sum_{i=1}^n X_i^2 + 2n\bar{X}_n \cdot \theta_0 - n\theta_0^2}{2}\right\} \\
&= \exp\left\{n\bar{X}_n(\bar{X}_n - \theta_0) - \frac{n(\bar{X}_n - \theta_0)(\bar{X}_n + \theta_0)}{2}\right\} \\
&= \exp\left\{\frac{n(\bar{X}_n - \theta_0)^2}{2}\right\}
\end{aligned}$$

对于 Rejection region, 其可以被简化为:

$$
\begin{aligned}
& T(X) > c \\
\Rightarrow \quad & \exp\left\{\frac{n(\bar{X}_n - \theta_0)^2}{2}\right\} > c \\
\Rightarrow \quad & \left|\bar{X}_n - \theta_0\right| > c'
\end{aligned}
$$

当 $\theta = \theta_1$ 时,

$$
\begin{aligned}
\beta(\theta_1) &= \mathbb{P}_{\theta_1}(\left|\bar{X}_n - \theta_0\right| > c') \\
&= \mathbb{P}_{\theta_1}(\bar{X}_n - \theta_0 > c') + \mathbb{P}_{\theta_1}(\bar{X}_n - \theta_0 < -c') \\
&= \mathbb{P}_{\theta_1}\left(\frac{\bar{X}_n - \theta_1}{\frac{1}{\sqrt{n}}} > \frac{c' + \theta_0 - \theta_1}{\frac{1}{\sqrt{n}}}\right) + \mathbb{P}_{\theta_1}\left(\frac{\bar{X}_n - \theta_1}{\frac{1}{\sqrt{n}}} < \frac{-c' + \theta_0 - \theta_1}{\frac{1}{\sqrt{n}}}\right) \\
&= 1 - \Phi(\sqrt{n} \cdot (c' + \theta_0 - \theta_1)) + \Phi(\sqrt{n} \cdot (-c' + \theta_0 - \theta_1))
\end{aligned}
$$

令 Type I error 为 α, 则有:

$$
\begin{aligned}
& \beta(\theta_0) = 2\Phi(-c' \cdot \sqrt{n}) = \alpha \\
\Rightarrow \quad & c' = \frac{z_{1-\alpha/2}}{\sqrt{n}}
\end{aligned}
$$

2.2 Theorem: Wilk's theorem

Logic

关于 Wilk's theorem 的详细描述, 见 STA3020 Lecture 10

考虑 hypothesis: $H_0 : \theta \in \Theta_0 \quad v.s. \quad H_1 : \theta \in \Theta_1$
令:

1. LRT 的 test statistic 为 $T(X)$
2. Θ 的 dimension 为 $rank(\Theta) = rank(\Theta_0 \cup \Theta_1) = r$, Θ_0 的 dimension 为 $rank(\Theta_0) = q$

若**在 Θ 和 Θ_1 之间存在 dimension difference**, 即 $r - q > 0$

则 as sample size $n \to \infty$, 有

$$
2\{\log(T(X))\} \xrightarrow{d} \chi^2_{r-q}
$$

Proof

Wilk's theorem 的详细证明见 STA3020 Lecture 10

Remark: dimension of parameter space

关于 dimension of parameter space, 我们需要考虑所有 unknown parameters,

例如, 考虑 $\mathcal{N}(\mu, \sigma^2)$, 其中 σ unknown, 则

1. $H_0 : \mu = 0, \sigma = 1, \quad v.s. \quad H_1 : \mu \neq 0, \sigma \neq 1 \quad \Rightarrow \quad dim(\Theta_0) = 0, \quad dim(\Theta) = 2$
2. $H_0 : \mu = 0, \quad v.s. \quad H_1 : \mu \neq 0 \quad \Rightarrow \quad dim(\Theta_0) = 1, \quad dim(\Theta) = 2$

Remark: Chi-square distribution

The χ^2_k distribution 为 sum of squares of k independent standard normal random variables, 即, 若 $Z_1, \ldots, Z_k \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$, 则

$$
Y = \sum_{i=1}^{k} Z_i^2 \sim \chi^2_k
$$

Example

问题:

令:

1. $X_1, \ldots, X_n \overset{i.i.d.}{\sim} Poisson(\theta)$
2. θ 的 MLE 为 $\hat{\theta}_n = \sum_{i=1}^n \frac{X_i}{n}$
3. $H_0 : \theta = \theta_0 \quad v.s. \quad H_1 : \theta \neq \theta_0$

求 LRT statistic 及其 asymptotic distribution, 并构造 size α LRT

解答:

LRT statistic 为:

$$\begin{aligned}\lambda &= 2\log\frac{\mathcal{L}(\hat{\theta}_n)}{\mathcal{L}(\theta_0)} \\ &= 2\log\frac{e^{-n\hat{\theta}_n}\hat{\theta}_n^{\sum_{i=1}^n X_i}}{e^{-n\theta_0}\theta_0^{\sum_{i=1}^n X_i}} \\ &= 2n[(\theta_0 - \hat{\theta}_n) - \hat{\theta}_n \log(\theta_0/\hat{\theta}_n)]\end{aligned}$$

对 $\log(\theta_0/\hat{\theta}_n)$ 做 Taylor expansion, 有:

$$\log(\theta_0/\hat{\theta}_n) = \log(\theta_0) - \log(\hat{\theta}_n) \approx \frac{1}{\hat{\theta}_n}(\theta_0 - \hat{\theta}_n) - \frac{1}{2\hat{\theta}_n^2}(\theta_0 - \hat{\theta}_n)^2$$

因此,

$$\lambda \approx 2n\left[(\theta_0 - \hat{\theta}_n) - (\theta_0 - \hat{\theta}_n) + \frac{1}{2\hat{\theta}_n}(\theta_0 - \hat{\theta}_n)^2\right] = \left(\frac{\theta_0 - \hat{\theta}_n}{\sqrt{\frac{\hat{\theta}_n}{n}}}\right)^2 \xrightarrow{d} \chi_1^2$$

因此 reject H_0 若 $\lambda > \chi_{1,\alpha}^2$, 其中 $\chi_{1,\alpha}^2$ 为 χ_1^2 的 α upper quantile