STAT 620: Asymptotic Statistics **Spring 2022**

Lecture: Feb 17

Lecturer: Xianyang Zhang

1 Testing and Confidence Set

Scientific method:

- Propose hypothesis
- Design experiment and observe data
- Reject the hypothesis or hypothesis remains consistent with the observed data

2 Confidence set

Goal: We want to find $\mathcal{C} \subset \mathbb{R}^d$ such that $P(\theta_0 \in \mathcal{C})$ has high probability.

2.1 Example

Suppose $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I_{\theta_0}^{-1})$ and I_θ is continuous of θ and invertible. Let

$$l_{n,r} = \{\theta \in \mathbb{R}^d : n(\hat{\theta}_n - \theta)^\top I_{\hat{\theta}_n}(\hat{\theta_n} - \theta) \le r\}.$$

Set $W \sim \mathcal{N}(0, I_{\theta_0}^{-1})$ and $Z \sim N(0, I_d)$. Notice

$$\begin{aligned} n(\hat{\theta}_n - \theta_0)^T I_{\hat{\theta}_n}(\hat{\theta}_n - \theta_0) &= \sqrt{n}(\hat{\theta}_n - \theta_0)^\top \left(I_{\theta_0} + o_p(1)\right) \sqrt{n}(\hat{\theta}_n - \theta_0) \\ &\xrightarrow{d} W^\top I_{\theta_0} W \\ &= Z^\top Z \sim \chi^2_d. \end{aligned}$$

Thus

$$P(\theta_0 \in l_{n,r}) \to P(\chi^2_d \le r).$$

2.2 General case

In general, we only need

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \Sigma) \quad \text{and} \quad \widehat{\Sigma} \xrightarrow{P} \Sigma.$$

Then we can construct confidence set in a similar way

$$\left\{\theta \in \mathbb{R}^d : n(\hat{\theta}_n - \theta)^\top \widehat{\Sigma}^{-1}(\hat{\theta}_n - \theta) \le r\right\}.$$

3 Testing: Dual problem to confidence sets

Definition:

Let $H_0 : \theta \in \Theta_0$, the *p-value* associated with a sample of observed values $x_1, \cdots, x_n$ is defined to be

$$\sup_{\theta \in \Theta_0} P_\theta(\text{data as extreme as } x_1, \cdots, x_n).$$

Example:

Let $H_0 : \theta = 0$. Suppose $X_i \sim^{i.i.d} N(\theta, 1)$ with the observed values $x_1, \cdots, x_n$. Then the p-value is

$$P_0(|\bar{Z}| > |\bar{x}|),$$

where $Z_i \sim^{i.i.d} N(0,1)$ for $i = 1, 2, \ldots, n$, $\bar{Z} = \frac{1}{n}\sum_{i=1}^n Z_i$ and $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$

3.1 Likelihood ratio test

Consider the following testing problem

$$H_0 : P = P_0 \quad \text{versus} \quad H_1 : P = P_1.$$

Let p_i be the pdf or pmf for P_i. The test that maximizes power at α level is the likelihood ratio test, which is given by

$$T(X_1, \ldots, X_n) = \log \frac{\prod_{i=1}^n p_1(X_i)}{\prod_{i=1}^n p_0(X_i)}.$$

Define

$$\phi(X_1, \ldots, X_n) = \begin{cases} 1 & \text{if } T(X_1, \ldots, X_n) > t, \\ 0 & \text{if } T(X_1, \ldots, X_n) < t, \\ r & \text{if } T(X_1, \ldots, X_n) = t, \end{cases}$$

where we pick t, r such that $\mathbb{E}_{P_0}[\phi(X_1, \ldots, X_n)] = \alpha$. In a more general case,

$$H_0 : \theta \in \Theta_0 \text{ versus } H_1 : \theta \in \Theta \backslash \Theta_0,$$

where $\Theta_0 \subset \Theta \subset \mathbb{R}^d$. We consider the test statistic

$$T(X_1, \ldots, X_n) = \log \left(\frac{\sup_{\theta \in \Theta} \prod_{i=1}^n p_\theta(X_i)}{\sup_{\theta \in \Theta_0} \prod_{i=1}^n p_\theta(X_i)} \right).$$

Definition:

For a sequence of test ϕ_n, the *uniform asymptotic level* of ϕ_n for $H_0 : \theta \in \Theta_0$ is

$$\limsup_{n \to \infty} \sup_{\theta \in \Theta_0} p_\theta(\phi_n \text{ reject } H_0).$$

The *pointwise level* for ϕ_n is

$$\sup_{\theta \in \Theta_0} \limsup_{n \to \infty} p_\theta(\phi_n \text{ reject } H_0).$$

It is easy to show that the pointwise level is less than the uniform asymptotic level.

3.2 Proposition

Let $\Theta_0 = \{\theta_0\}$ and $L_n(\theta) = \sum_{i=1}^n l_\theta(X_i) = \sum_{i=1}^n \log(p_\theta(X_i))$. Define

$$\Delta_n = \log\left(\frac{\max_{\theta\in\Theta}\prod_{i=1}^n p_\theta(X_i)}{\prod_{i=1}^n p_{\theta_0}(X_i)}\right) = \log\left(\frac{\prod_{i=1}^n p_{\hat\theta_n}(X_i)}{\prod_{i=1}^n p_{\theta_0}(X_i)}\right) = L_n(\hat\theta_n) - L_n(\theta_0).$$

Then under regularity conditions for proving the asymptotic efficiency for MLE,

$$2\Delta_n \xrightarrow[H_0]{d} \chi^2_d,$$

where d is the dimension of parameters.

Proof For large n,

$$0 = \nabla L_n(\hat\theta_n) = \nabla L_n(\theta_0) + \nabla^2 L_n(\theta_0)(\hat\theta_n - \theta_0) + \sum_{i=1}^n \hat r(x_i)(\hat\theta_n - \theta_0)$$

and from a previous lemma

$$\|P_n \hat r\|_{op} = o_p(1).$$

We take a second order Taylor expansion for L_n,

$$L_n(\hat\theta_n) = L_n(\theta_0) + \nabla L_n(\theta_0)(\hat\theta_n - \theta_0) + \frac{1}{2}(\hat\theta_n - \theta_0)^\top \nabla^2 L_n(\theta_0)(\hat\theta_n - \theta_0) + o_p(1).$$

Therefore,

$$\begin{aligned}
\Delta_n &= \nabla L_n(\theta_0)(\hat\theta_n - \theta_0) + \frac{1}{2}(\hat\theta_n - \theta_0)^\top \nabla^2 L_n(\theta_0)(\hat\theta_n - \theta_0) + o_p(1) \\
&= -\frac{1}{2}(\hat\theta_n - \theta_0)^\top \nabla^2 L_n(\theta_0)(\hat\theta_n - \theta_0) - n(\hat\theta_n - \theta_0)^T P_n \hat r(\hat\theta_n - \theta_0) + o_p(1) \\
&= -\frac{1}{2}(\hat\theta_n - \theta_0)^\top \nabla^2 L_n(\theta_0)(\hat\theta_n - \theta_0) + o_p(1).
\end{aligned}$$

Thus

$$\begin{aligned}
2\Delta_n &= \sqrt{n}(\hat\theta_n - \theta_0)^T\left(-\frac{\nabla^2 L_n(\theta_0)}{n}\right)\sqrt{n}(\hat\theta_n - \theta_0) + o_p(1) \\
&\xrightarrow{d} W^\top I_{\theta_0} W = \chi^2_d,
\end{aligned}$$

where $W \sim N(0, I_{\theta_0}^{-1})$.