STAT 620: Asymptotic Statistics **Spring 2022**

Lecture: Feb 22

Lecturer: Xianyang Zhang

1 Wald's test

A Wald confidence eclipse

$$C_{n,\alpha} = \left\{ \theta \in \mathbb{R}^d : (\theta - \hat{\theta}_n) I_{\hat{\theta}_n} (\theta - \hat{\theta}_n) \leq \frac{r_{d,\alpha}}{n} \right\},$$

where $r_{d,\alpha}$ is determined by $P(\chi_d^2 \geq r_{d,\alpha}) = \alpha$.

A Wald's test of a point null hypothesis $\theta = \theta_0$ is

$$T_n = \begin{cases} \text{reject}, & \theta_0 \notin C_{n,\alpha}, \\ \text{fail to reject}, & \theta_0 \in C_{n,\alpha}. \end{cases}$$

That is we reject the null hypothesis if $n(\hat{\theta}_n - \theta_0) I_{\hat{\theta}_n} (\hat{\theta}_n - \theta_0) \geq r_{d,\alpha}$. One can also replace $I_{\hat{\theta}_n}$ by I_{θ_0}.

2 Testing a subvector of θ

2.1 Schur complement and Wald test

Suppose $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$, $A = A^\top$, and $A \succ 0$. Then,

$$A^{-1} = \begin{bmatrix} (A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1} & * \\ * & * \end{bmatrix}.$$

Assume that $\sqrt{n}(\hat{\theta}_n - \theta_0) \to^d N(0, \Sigma_{\theta_0})$, where $\Sigma_{\theta_0} = I_{\theta_0}^{-1}$. Define

$$[v]_{1:k} = \begin{bmatrix} v_1 \\ \vdots \\ v_k \end{bmatrix}$$

for $v \in \mathbb{R}^d$ and $k \leq d$. For $\Sigma \in \mathbb{R}^{d \times d}$, let

$$\Sigma = \begin{bmatrix} \Sigma^{(k)} & * \\ & * \end{bmatrix},$$

where $\Sigma^{(k)} \in \mathbb{R}^{k \times k}$. Then we have

$$\sqrt{n}([\hat{\theta}_n]_{1:k} - [\theta_0]_{1:k}) \to N(0, \Sigma_{\theta_0}^k)$$

where $\Sigma_{\theta_0}^k = (I_{11,\theta_0} - I_{12,\theta_0} I_{22,\theta_0}^{-1} I_{21,\theta_0})^{-1}$. Consider testing $H_0 : [\theta]_{1:k} = [\theta_0]_{1:k}$. The Wald's test is given by

$$n([\hat{\theta}_n]_{1:k} - [\theta_0]_{1:k})(\Sigma_{\hat{\theta}_n}^{(k)})^{-1}([\hat{\theta}_n]_{1:k} - [\theta_0]_{1:k}),$$

where $\hat{\theta}_n = [\theta_{01}, \ldots, \theta_{0k}, \tilde{\theta}_{k+1}, \ldots, \tilde{\theta}_n]$. Here $[\tilde{\theta}_{k+1}, \ldots, \tilde{\theta}_n]$ only needs to be asymptotically consistent (not necessarily efficient).

2.2 Likelihood ratio-test for subvector

The likelihood ratio test can be defined as

$$\log\left(\frac{\sup_{\theta\in\Theta}\Pi_{i=1}^n p_\theta(X_i)}{\sup_{\theta\in\Theta_0}\Pi_{i=1}^n p_\theta(X_i)}\right),$$

where $\Theta_0 = \{\theta \in \mathbb{R}^d : [\theta]_{1:k} = [\theta_{01}, \ldots, \theta_{0k}]\}$.

3 Rao-Score Test

Since $P_n \nabla l_\theta = \frac{1}{n}\sum_{i=1}^n \nabla l_\theta(X_i)$ and $P_\theta \nabla l_\theta = 0$, we have

$$\sqrt{n} P_n \nabla l_\theta \xrightarrow{d} N(0, I_\theta),$$

when the true parameter is equal to θ. Suppose we want to test a point null $H_0 : \theta = \theta_0$. Note that under H_0

$$n P_n \nabla l_{\theta_0} I_{\theta_0}^{-1} P_n \nabla l_{\theta_0} \to \chi_d^2.$$

Hence we reject H_0 if

$$n P_n \nabla l_{\theta_0} I_{\theta_0}^{-1} P_n \nabla l_{\theta_0} > r_{d,\alpha}.$$

4 U-statistic

4.1 Motivation

Suppose $h : \mathcal{X}^r \to \mathbb{R}$ and we want to estimate $\theta = E[h(X_1, \ldots, X_r)]$, where X_i's are i.i.d. random variables from some distribution. Natural questions to raise

- how to estimate θ?
- how to perform inference on θ, i.e. hypothesis testing or confidence interval construction?

For example, if $h(x_1, x_2) = \frac{1}{2}(X_1 - X_2)^2$, then $E[h(X_1, X_2)] = \text{var}(X_1)$.

4.2 Definition

For $X_i \sim^{i.i.d} P$, denote $\theta(P) = E_P[h(X_1, \ldots, X_r)]$. A U-statistic is random variable of the form

$$u_n = \frac{1}{\binom{n}{r}} \sum_{|\beta|=r, \beta\subset[n]} h(X_\beta),$$

where $[n] = \{1, \ldots, n\}$ and $X_\beta = \{X_\tau, \tau \in \beta\}$, h is symmetric (permutation of X_β gives the same value).

4.3 Remark

One way to construct symmetric function is by considering

$$\tilde{h}(x_1, \ldots, x_r) = \frac{1}{r!}\sum_\pi h(x_{\pi(1)}, \ldots, x_{\pi(r)}),$$

where the summation is over all permutations of $\{1, 2, \ldots, r\}$.

4.4 V-statistic

A V-statistic for estimating θ is given by

$$V_n = \frac{1}{n^r} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_r=1}^{n} h(X_{i_1}, \ldots, X_{i_r}).$$

Note that V_n is biased (since $i_1, \ldots, i_r$ can be equal).

4.5 Theorem

For any convex loss function $L(\cdot)$ and random variable Z_n such that $E(Z_n | X_{(i)}, 1 \leq i \leq n) = u_n$, we obtain that

$$E(L(Z_n)) \geq E(L(u_n)).$$

Proof Note that

$$E(L(u_n)) = E(L(E(Z_n | X_{(i)}, 1 \leq i \leq n))) \leq E(E(L(Z_n) | X_{(i)}, 1 \leq i \leq n)) = E(L(Z_n)),$$

where we have used the Jensen's inequality.

4.6 Example

Let $\{X_{(1)}, \ldots, X_{(n)}\}$ be the order statistics of $\{X_1, \ldots, X_n\}$. We know that

$$E[h(X_1, \ldots, X_r) | X_{(1)}, \ldots, X_{(n)}] = u_n.$$