Stats 300b: Theory of Statistics	Winter 2017

Lecture 9 – February 7

Lecturer: John Duchi *Scribe: Dylan Cable*

Warning: *these notes may contain factual errors*

Reading: Chapter 7.7 of Lehmann's Elements of Large Sample Theory

Today we are going to finish up hypothesis and confidence tests. We will return to them when we talk about optimality. Then, we will talk about uniform laws of large numbers - how you give law of large numbers uniformly over a function class.

Outline :
Parametric Big 3 tests (Chapter 7.7 of Lehmann's Elements of Large Sample Theory)

- Likelihood Ratio Test
- Wald Test
- Rao (score tests)

Begin Uniform Laws of Large Numbers (ULLNs)

- Concentration inequalities
- Sub-Gaussianity

Recap: Generalized Likelihood Ratio Test: Let $(x_1, \ldots, x_n) = x$, $L_n(x;\theta) := \sum_{i=1}^n l_\theta(x_i), l_\theta(x) = \log p_\theta(x)$

Let $\hat{\theta}_n = \text{argmax}_\theta L_n(x;\theta)$ and assume usual asymptotic normality conditions hold (twice differentiable, Lipschitz-continuous hessian). Then $\delta_n = L_n(x;\hat{\theta}_n) - L_n(x;\theta)$ satisfies

$$2\delta_n \underset{P_{\theta_0}}{\to} \chi^2_d, \theta_0 \in \mathbb{R}^d$$

Definition: Let T_n be a sequence of tests for some model $\{P_\theta\}_{\theta\in\Theta}$, let

$$H_0 : \theta \in \Theta_0 \subset \Theta$$

Then T_n is <u>asymptotically level α</u> if

$$\lim_{n\to\infty} \sup_{\theta\in\Theta_0} P_\theta(T_n \text{ rejects } H_0) \le \alpha$$

Example: Generalized Likelihood Ratio Test is defined by reject if $\delta_n \ge u^2_{d,\alpha}/2$, where $u^2_{d,\alpha}$ is defined so that $\mathbb{P}(\chi^2_d \ge u^2_{d,\alpha}) = \alpha$. This test is asymptotically level α.

<u>Wald Tests</u> Define <u>Wald confidence ellipsoid</u> for model $\{P_\theta\}$ with Fisher information I_θ by

$$C_{n,\gamma} = \{\theta \in \Theta : (\theta - \hat{\theta}_n)^T I_{\hat{\theta}_n}(\theta - \hat{\theta}_n) \le \gamma/n\}$$

where $\hat{\theta}_n$ is Maximum Likelihood Estimator. We saw last time that

$$n(\theta - \hat{\theta}_n)^T I_{\hat{\theta}_n}(\theta - \hat{\theta}_n) \underset{P_{\theta_0}}{\to} \chi^2_d$$

Wald test of $H_0 : \theta = \theta_0$ against $\theta \neq \theta_0$:

Acceptance: $\hat{\theta}_n \in \{\theta \in \Theta : (\theta - \theta_0)^T I_{\hat{\theta}_n}(\theta - \theta_0 \leq u^2_{d,\alpha}/n\} := A_n$

This is asymptotically level α. Why?

$$n(\hat{\theta}_n - \theta_0)^T I_{\hat{\theta}_n}(\hat{\theta}_n - \theta_0) \underset{P_{\theta_0}}{\to} \chi^2_d$$

$$P_{\theta_0}(\hat{\theta}_n \notin A_n) = P_{\theta_0}((\hat{\theta}_n - \theta_0)^T I_{\hat{\theta}_n}(\hat{\theta}_n - \theta_0) > u^2_{d,\alpha}/n) \to \mathbb{P}(\chi^2_d \geq u^2_{d,\alpha}) = \alpha$$

Remark: If $H_0 = \theta = \theta_0$ is a point null, we can replace $I_{\hat{\theta}_n}$ with I_{θ_0}.

What about composite nulls (nuisance parameters)?

Example:

$X_i \underset{i.i.d.}{\sim} N(\theta, \sigma^2), \sigma^2$ is unknown.

$H_0 : \theta = 0$, but σ^2 is unspecified.

To deal with this, let $\Sigma(\theta) = I^{-1}(\theta)$ so that $\sqrt{n}(\hat{\theta}_n - \theta) \underset{P_{\theta_0}}{\overset{d}{\to}} N(0, \Sigma(\theta))$

For vector $v \in \mathbb{R}^d$, let $[v]_{1:k} = \begin{bmatrix} v_1 \\ \vdots \\ v_k \end{bmatrix} \in \mathbb{R}^k$

For $\Sigma = (e_{ij})^d_{i,j=1} \in \mathbb{R}^{d \times d}$, let $\Sigma^{(k)}$ be the leading k-by-k principal submatrix $(e_{ij})^k_{i,j=1}$

Then $\sqrt{n}([\hat{\theta_n}]_{1:k} - [\theta]_{1:k}) \to N(0, \Sigma^{(k)}(\theta))$

Fact: if $A = \begin{bmatrix} A_{11} A_{12} \\ A_{21} A_{22} \end{bmatrix}, M = A^{-1}$, then $M_{11} = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}$ (exercise).

Then $\Sigma^{(k)}(\theta) = ((I(\theta))^{-1})^k = (I^k(\theta) - I_{12}(\theta)I_{22}(\theta)^{-1}I_{21}(\theta))^{-1}$

Then $\Sigma^{(k)}(\theta) \succeq (I^{(k)}(\theta))^{-1}$ and equal if and only if $I_{12}(\theta) = 0$ (i.e. $[\hat{\theta_n}]_{1:k}$ and $[\hat{\theta_n}]_{k+1:d}$ are asymptotically independent.)

$$n([\hat{\theta_n}]_{1:k} - [\theta_0]_{1:k})^T \Sigma^{(k)}(\hat{\theta_n})^{-1}([\hat{\theta_n}]_{1:k} - [\theta_0]_{1:k}) \underset{\theta_0}{\to} \chi^2_k$$

Let $H_0 : \theta_1 = \theta^0_1, \ldots, \theta_k = \theta^0_k$, and $\theta_{k+1}, \ldots, \theta_d$ are unspecified

$$\hat{\theta_n} \in \{\theta : ([\theta]_{1:k} - [\theta^0]_{1:k})^T \Sigma^{(k)}(\hat{\theta_n})^{-1}([\theta]_{1:k} - [\theta^0]_{1:k}) \leq u^2_{k,\alpha}/n\}, \mathbb{P}(\chi^2_k \geq u^2_{k,\alpha}) = \alpha$$

Then this test is asymptotically level α. We must use $\Sigma(\hat{\theta_n}) = I(\hat{\theta_n})^{-1}$ to get a consistent estimator of $I(\theta)$ under H.

Example: Gaussian mean, unknown covariance. $N(\theta, \Sigma), \theta \in \mathbb{R}^d, \Sigma \succ 0$

$\bar{x}_n = \frac{1}{n}\sum^n_{i=1} x_i$, $\hat{\Sigma} = \frac{1}{n}\sum^n_{i=1}(x_i - \bar{x}_n)(x_i - \bar{x}_n)^T$

The test is reject if $\{\bar{x}^T_n \hat{\Sigma}^{-1}\bar{x}_n \geq u^2_{d,\alpha})/n\}$ This is because the fisher information of Gaussian is $\hat{\Sigma}^{-1}$.

Rao Score Test: Why the Rao Score Test? The MLE is a bit difficult to compute.

Suppose I have a point null $H_0 : \theta = \theta_0$

What is the limit of $\sqrt{n}P_n\nabla l_{\theta_0}$ under H_0?

$$\sqrt{n}P_n\nabla l_{\theta_0} \to N(0, I_{\theta_0})$$

$$n(P_n \nabla l_{\theta_0})^T I_{\theta_0}^{-1} (P_n \nabla l_{\theta_0}) \underset{P_{\theta_0}}{\to} \chi^2_d$$

Define: Rao test of $H_0 : \theta = \theta_0$ vs $\theta \neq \theta_0$: Reject if $(P_n \nabla l_{\theta_0})^T I_{\theta_0}^{-1} (P_n \nabla l_{\theta_0}) \geq u^2_{d,\alpha}/n$. This is asymptotically level α.

Remarks: (1) Analogues for composite nulls. (2) Strong connections to optimal testing (we will see later).

Uniform Laws of Large Numbers: (ULLN) Basic questions: Given collection $\mathcal{F}$ of functions $\subset \{f : X \to \mathbb{R}\}$, when do we have $\sup_{f \in \mathcal{F}} |P_n f - Pf| \underset{p}{\to} 0$?

Why should we care? M-estimators: Let $m_\theta : X \to \mathbb{R}$ be defined for $\theta \in \Theta$. The associated M-estimator is

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmax}}\, P_n m_\theta(x)$$

When does $\hat{\theta}_n \to \operatorname{argmax}_{\theta \in \Theta} Pm_\theta(x)$

Let $M(\theta) := Pm_\theta(x)$ be the population objective. $M_n(\theta) := P_n m_\theta(x)$.

When does $M(\hat{\theta}_n) \underset{p}{\to} \sup_{\theta \in \Theta} M(\theta)$?

Suppose we had a ULLN: $\sup_{\theta \in \Theta} |M_n(\theta) - M(\theta)| \underset{p}{\to} 0$

Let $\hat{\theta}_n = \operatorname{argmax}_{\theta \in \Theta} M_n(\theta)$

$M_n(\hat{\theta}_n) = \frac{1}{n} \sum_{i=1}^n m_{\hat{\theta}_n}(x_i)$ is not an i.i.d sum because $\hat{\theta}_n$ depends on (x_i).

But $M(\hat{\theta}_n) - M(\theta^*) = (M(\hat{\theta}_n) - M_n(\hat{\theta}_n)) + (M_n(\hat{\theta}_n) - M_n(\theta^*)) + (M_n(\theta^*) - M(\theta^*))$

$\geq \sup_{\theta \in \Theta} |M_n(\theta) - M(\theta)| + 0 + o_p(1)$

Applying our ULLN, we conclude $M(\hat{\theta}_n) \underset{p}{\to} M(\theta^*)$