STAT5010 Advanced Statistical Inference **Fall 2023**

Lecture 9: Least Favorable Distribution and Asymptotic Optimality

Lecturer: Tony Sit *Scribe: Ji Qi, Zewu Zheng*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

9.1 P-value

Suppose we want to test $H_0 : \theta \leq 0$ vs. $H_1 : \theta > 0$ at level $0 < \alpha < 1$. Here $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\theta, 1)$. The UMP test is

$$\phi_\alpha = \begin{cases} 1 & \sum X_i > Z_{1-\alpha}\sqrt{n} \\ 0 & otherwise \end{cases}$$

Let $S_\alpha = \{X : \sum_{i=1}^n X_i > Z_{1-\alpha}\sqrt{n}\}$ be the secretion region. If $\alpha_1 < \alpha_2$, then $S_{\alpha_1} \subseteq S_{\alpha_2}$.

Suppose we want to test H_0 vs H_1 at level α. Let ϕ_α test at level α. Assume that the rejection regions an nested, i.e. $\alpha_1 < \alpha_2 \Rightarrow S_{\alpha_1} \subseteq S_{\alpha_2}$ where $S_\alpha = \{x : \phi_\alpha(x) = 1\}$.

Definition 9.1 $\hat{p}(x) = \inf\{u : x_\in \mathrm{S_u}\}$.

Intuitively, given the p-value, you can construct a level α-test by rejecting H_0 if $\hat{p}(x) < \alpha$ and accepting H_0 if $\hat{p}(x) > \alpha$, e.g.

$$\begin{aligned} S_\alpha &= \left\{X_i \sum X_i > Z_{1-\alpha}\sqrt{n}\right\} \\ &= \left\{X_{:}1 - \Phi\left(\frac{\Sigma X_i}{\sqrt{n}}\right) < \alpha\right\} \\ &\Rightarrow \hat{p}(x) = 1 - \Phi\left(\frac{\Sigma x_i}{\sqrt{h}}\right) \end{aligned}$$

under $H_0 : g = 0, \hat{p}(x) \sim U(0,1)$

$$P(\hat{p}(x) \leq u) = u$$

Lemma 9.2 *Suppose $X \sim P_\theta$ for some $\theta \in \Theta$. We want to test $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$ at level α. Let $\{\phi_\alpha\}_{\alpha \in (0,1)}$ be a collection of nested level α tests. Then*
(i) $P_\theta(\hat{p}(x) \leq u) \leq u = P(u(0,1) \leq u), \forall u \in (0,1), \theta \in \Theta_0$
(ii) If $\exists g_0 \in \Theta_0$, such that $P_{\theta_0}(X \in S_\alpha) = \alpha$ for $\forall \alpha$, then $P_{\Theta_0}(\hat{p}(x) \leq u) = u$.

Definition 9.3 ***(Confidence interval)*** *Let $X \sim P_\theta$ for some $\theta \in \Theta$. For every $X \in X$, Let $S(X)$ be a subset of Θ. We say the collection of sets $\{S(X), X \in \mathcal{X}\}$ is an $(1-\alpha)$ confidence region if $P_\theta(\theta \in S(x)) \geq 1 - \alpha, \forall \theta \in \Theta$.*

9.2 Asymptotic Optimality

Let $\{X_1, \cdots, X_n\}$ be i.i.d. from $\{P_\theta, 0 \in \Theta\}$ with pdf w.r.t. some σ-finite measure. Suppose we want to estimate $g(\theta)$, and a candidate estimator is $\delta_n(x_1, \ldots, x_n)$.

Definition 9.4 *We say $\delta_n(x)$ is consistent for $g(\theta)$, if $\delta_n(x) \xrightarrow{p} g(\theta), \quad \forall \theta \in \theta$, i.e. $\forall \theta \in \Theta, \forall \varepsilon > 0$, we have $P_\theta(|\delta_n(X) - g(\theta)| > \epsilon) \to 0$.*

Remark: If $X_1 \ldots X_n \overset{\text{iid}}{\sim}$ F:
(i)Assume $E_F|X| < \infty$, then $\frac{1}{n}\sum x_i \xrightarrow{p} E_F(X)$ (WLLN).
(ii) Assume $E_F X^2 < \infty$, then $W_n = \frac{\sum X_i - nE_p(X)}{\sqrt{n \operatorname{var_p}(X)}} \xrightarrow{d} \mathcal{N}(0,1)$(CLT),

$$\Leftrightarrow \lim_{n\to\infty} P(W_n \leq t) = \Phi(t), \quad \forall t \in R$$

(Example) $X_1 \ldots X_n \overset{\text{iid}}{\sim}$ Bernoulli (θ), if $\theta \in Q$; $X_1 \ldots X_n \overset{\text{iid}}{\sim}$ Bernoulli $(1-\theta)$. if $\theta \notin Q$; then there is no consistent estimator of $X_1 \ldots X_n$.

Definition 9.5 *Let $L(\theta|x) = \prod_{i=1}^n p_\theta(x_i)$ be the likelihood function. If there exists a unique θ_n which is the global maximizer of $\theta \to L(\theta/x)$(or $\theta \to l(\theta \mid x) = \log L(\theta \mid x)$). Define $\hat{\theta}_n$ as the MLE of θ.*

(Example) $X_1 \ldots X_n \overset{\text{iid}}{\sim}$ Bernoulli (θ). $\log(\theta|x) = \sum_{i=1}^n x_i(\log\theta) + (n - \sum_{i=1}^n x_i)\log(1-\theta)$,
$\Rightarrow \hat{\theta}_n = \arg\max_\theta \ln(\theta \mid x) = \frac{\sum_{i=1}^n x_i}{n} = \sum_{i=1}^n x_i/n$.
(i) $\bar{X}_n \xrightarrow{p} \theta, \quad \forall \theta \in (0,1)$, (consistency)
(ii) $\sqrt{n}(\overline{x_n} - \theta) \xrightarrow{d} N\left(0. \frac{1}{\theta(1-\theta)}\right)$, (CLT).

Theorem 9.6 *Suppose $X_1 \ldots X_n \overset{iid}{\sim} P_\theta$ for some $\theta \in \Theta$ with pdf $P_\theta(\cdot)$
Assume A0: $P_{\theta_1} \neq P_{\theta_2}$. where $\theta_1 \neq \theta_2$ [identifiability]; A1: $\{P_\theta(\cdot), \theta \in \Theta\}$ has common support.
Then we have: $P_{\theta_0}(\log(\hat{\theta n} \mid x) > \log(\theta \mid x)) \to 1$ as $n \to \infty, \quad \forall \theta \neq \theta_0$.*

Proof: Let $T_n = \frac{1}{n}\sum_{i=1}^n \log\frac{p_\theta(x_i)}{p_\theta(x_i)}$, then $T_n \xrightarrow{p} E_{\theta_0}\log\frac{p_\theta(x_1)}{p_{\theta_0}(x_2)}$. Now $E_\theta \log\frac{P_\theta(X_1)}{P_{\theta_0}(X_1)} = \int \log\frac{p_0(x)}{p_\theta(x)} p_\theta(x) d\mu(x) = -D(\theta_0||\theta) < 0$ for $\theta \neq \theta_0$. Hence, $P_{\theta_0}(T_n \neq 0) \to 1$, but $T_n < 0 \Leftrightarrow \frac{1}{n}\sum_{i=1}^n \log\frac{P_\theta(X_i)}{P_\theta(X_i)} < 0 \Leftrightarrow \log\prod_{i=1}^n P_\theta(X_i) < \log\prod_{i=1}^n P_{\theta_0}(X_i) \Leftrightarrow \ell_n(\theta|x) < \ell_n(\theta|x)$. ∎

Corollary 9.7 *Suppose A0 and A1 hold, it Θ is finite, then the MLE $\hat{\theta_n}$ exists with high probability and $P_{\theta_0}(\hat{\theta_n} = 0) \to 1, \quad (n \to \infty)$.*

Suppose $A0$ and $A1$ hold. Suppose that $\Theta \subseteq \mathbb{R}$ and θ_0 is an interior point of Θ. If $\theta \mapsto p_0(x)$ is differentiable and the deviates is continuous. there exist a sequence of roots $\hat{e}_n$ of the likelihood equation $\frac{\partial}{d\theta}\operatorname{In}(\theta/x) = 0$. where is consistent for θ_0.

Let $A_n = \{x : \ln(\theta_0 \mid x) > \max_{j \leq k} \ln(\theta_j \mid x)\}$. If $X \in A_n$, then $\hat{\theta_n}(x) = \theta_0$ and $P_{\theta_0}(A_n) \to 1$.

Theorem 9.8 *Suppose A0 and A1 hold. Suppose further that A2: $\Theta \subset \mathbb{R}$ and θ_0 is an interior point of Θ. If $\theta \mapsto P_\theta()$ is differentiable and the derivative is continuous, then there exists a sequence of roots $\hat{\theta}_n$ of the score function $\ell_n'(\theta)\partial\ell_n(\theta\,|)/\partial\theta = 0$, which is consistent for θ_0.*

Proof: Let $\delta > 0$ be small enough such that $[\theta_0 - \delta, \theta_0 + \delta] \subset \Theta$. It follows that

$$P_{\theta_0}\left(\ell_n(\theta_0 \mid) > \ell_n(\theta_0 \pm \delta \mid)\right) \to 1$$

as $n \to \infty$. Now, the function $\theta \mapsto \ell_n(\theta \mid)$ is a continuous function on the compact set $[\theta_0 - \delta, \theta_0 + \delta]$. There exists a global maximiser $\tilde{\theta}_n(\delta)$. But $\tilde{\theta}_n(\delta)$ cannot be $\theta_0 \pm \delta$ as θ_0 is better, which implies that $\ell_n'(\tilde{\theta}_n(\delta)) = 0$.

Let $\hat{\theta}_n(\delta)$ denote the closest root of $\ell_n'(\theta) = 0$ to θ_0. Fix $\delta > 0$, we need to show that $P_{\theta_0}(|\hat{\theta}_n - \theta_0| < \delta) \to 1$ as $n \to \infty$. Observe that $|\hat{\theta}_n - \theta_0| \leq |\tilde{\theta}_n(\delta) - \theta_0|$ as $\hat{\theta}_n$ is the closet root. It follows that

$$P_{\theta_0}(|\hat{\theta}_n - \theta_0| < \delta) \geq P_{\theta_0}(|\tilde{\theta}_n(\delta) - \theta_0| < \delta) \geq P_{\theta_0}(\ell_n(\theta_0) > \ell_n(\theta_0 \pm \delta)) \to 1.$$

It remains to prove that there exists a closest root, i.e. $\exists \hat{\theta}$ such that $f(\hat{\theta}) = 0$, $|\hat{\theta} - \theta_0| = \inf_{\tilde{\theta}: f(\tilde{\theta})=0} |\tilde{\theta} - \theta_0|$, assuming that $\{\tilde{\theta} : f(\tilde{\theta}) = 0\}$ is non-empty, and $f(\cdot)$ is a continuous function on $\mathbb{R}$. To see this, let $\alpha = \inf_{\tilde{\theta}: f(\tilde{\theta})=0} |\tilde{\theta} - \theta_0|$. For all $k \geq 1$, there exists $\tilde{\theta}_k$ such that

$$f(\tilde{\theta}_k) = 0 \quad \text{and} \quad |\tilde{\theta}_k - \theta_0| \leq \alpha + k^{-1} \leq \alpha + 1. \tag{9.1}$$

Note also that $\tilde{\theta}_k \in [\theta_0 - \alpha - 1, \theta_0 + \alpha + 1]$. By going to a subsequence, as $k \to \infty$, $\tilde{\theta}_k \to \hat{\theta}$, say. But $|\hat{\theta} - \theta_0| = \alpha$ by taking the limit on (9.1) and the fact that $f(\hat{\theta}) = 0$ since $f(\cdot)$ is continuous. ■

Corollary 9.9 *If A0–A2 hold, assume further that $\theta \mapsto P_\theta()$ is differentiable, and the score function $\ell_n'(\theta) = 0$ has a unique root $\hat{\theta}_n$, then $\hat{\theta}_n \xrightarrow{P} \theta_0$ (follows from the previous theorem), and $\hat{\theta}_n$ is the MLE with probability tending to 1.*

Proof: It follows from the previous proof that $\hat{\theta}_n$ is a local maximum (with high probability). If $\hat{\theta}_n$ is not the unique global mimimiser of $\theta \mapsto \ell_n(\theta)$, then there exists $\tilde{\theta}_n$ such that $\ell_n(\tilde{\theta}_n) \geq \ell_n(\hat{\theta}_n), \tilde{\theta} \neq \hat{\theta}_n$. Then there exists $\breve{\theta}$ such that $\ell_n(\breve{\theta}) = \ell_n(\hat{\theta}_n), \breve{\theta} \neq \hat{\theta}$, as $\theta \mapsto \ell_n(\theta)$ is continuous. It implies that there exists $\epsilon_n \neq \hat{\theta}_n$ such that $\ell_n'(\epsilon_n) = 0$ [see Rolle's Theorem], which is a contradiction. ■

Theorem 9.10 *(Slutsky's Theorem) Suppose $X_n \xrightarrow{d} X$, $A_n \xrightarrow{p} a$, $B_n \xrightarrow{p} b$, then $A_n X_n + B_n \xrightarrow{d} aX + b$*

$$\begin{aligned}
0 = \ln'\left(\hat{\theta}_n\right) &= \ln'(\theta_0) + \left(\hat{\theta}_n - \theta_0\right)\ln_n''(\theta_0) + \frac{1}{2}\left(\hat{\theta}_n - \theta_0\right)^2 \ln'''(\xi_n) \\
\Rightarrow \left(\hat{\theta}_n - \theta_0\right)\left(l_n''(\theta_0) + \frac{1}{2}\left(\hat{\theta}_n - \theta_0\right)\ln'''(\xi_n)\right) &= -\ln'(\theta_0) \\
\Rightarrow \sqrt{n}\left(\hat{\theta}_n - \theta_0\right) &= \frac{-\ln'(\theta_0)/\sqrt{n}}{-l_n''(\theta_0)/n - \frac{1}{2}\left(\hat{\theta}_n - \theta_0\right)l_n'''(\xi_n)\Big)/n}
\end{aligned}$$

It suffices to show $\left.\begin{array}{r} \frac{1}{\sqrt{n}}\ln'(\theta_0) \xrightarrow{D} N(0, I(\theta_0)) \\ -\frac{1}{n}l_n''(\theta_0) \xrightarrow{P} I(\theta_0) \\ \text{and } \frac{1}{n}\left(\hat{\theta}_n - \theta_0\right)\ln(\xi_n) \xrightarrow{p} 0 \end{array}\right\}$, then $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} \mathcal{N}(0, I^{-1}(\theta_0))$