STAT 701 Lecture, February 12, 2025

The topics for this lecture are:

(1) Calculations for one-sided Likelihood Ratio Test

(2) View of LRT as generalization of Neyman-Pearson Lemma

(3) Consistency of MLE on finite Θ

(4) Generalization of MLE consistency to Compact Θ

Example with One-sided LRT-based Hypothesis Test

Consider $H_0 : \theta \le 2$ versus $H_A : \theta > 2$ based on data $X_i \sim f(x,\theta) = \theta\, x^{-\theta-1}\, I_{[x \ge 1]}$, where $\Theta = (0, \infty)$.

LRT statistic $\quad \Lambda = \frac{\max_{\theta \le 2} \; \theta^n \, (X_1 \cdots X_n)^{-\theta-1}}{\max_\theta \; \theta^n \, (X_1 \cdots X_n)^{-\theta-1}} \le 1$

Since $\log L(\theta; \underline{X}) = n \log\theta - (\theta+1) \sum_{i=1}^n \log X_i$ is strictly concave, calculus max $\widehat{\theta} = n / \sum_{i=1}^n \log X_i$ exists and is unique, but if $\widehat{\theta} > 2$, then $n/\theta - \sum_{i=1}^n \log X_i$ is positive and decreasing on $[0, 2]$, so max occurs at 2, and

$$\Lambda = I_{[\widehat{\theta} \le 2]} + I_{[\widehat{\theta} > 2]} \cdot e^n \, (X_1 \cdots X_n)^{-3} \, (2/\widehat{\theta})^n$$

LRT as generalization of Neyman-Pearson Lemma

If $\Theta = \{\theta_0, \theta_1\}$, then $\Lambda = L(\theta_0; \underline{X}) \,/\, \max\{L(\theta_0; \underline{X}), L(\theta_1; \underline{X})\}$

Test rejects for small values of Λ, i.e., $\Lambda \leq \lambda_\alpha < 1$ which means that $L(\theta_1; \underline{X})/L(\theta_0; \underline{X}) \geq 1/\lambda_\alpha$

Consistency of MLE on finite Θ

Suppose that the parameter space $\Theta = \{\theta_0, \theta_1, \ldots, \theta_k\}$ with data $(X_i,\ i = 1, \ldots, n)$ sampled from density $f(x, \theta)$ on $x \in \mathbb{R}^d$ for some $\theta \in \Theta$.

If $\theta = \theta_0$, then for each $j \geq 1$:

$$\frac{1}{n} \log \{ \prod_{i=1}^{n} \frac{f(X_i, \theta_j)}{f(X_i, \theta_0)} \} \longrightarrow E_{\theta_0}[\log(f(X_1, \theta_j)/f(X_1, \theta_0)]$$

$$\leq \log \{ E_{\theta_0}(f(X_1, \theta_j)/f(X_1, \theta_0)) \} = 0$$

Why is this true without assumptions on the densities ?

Jensen's Inequality $\Rightarrow$ Information Inequality

The reason is that $Z_i = f(X_i, \theta_j)/f(X_i, \theta_0)$ is a nonnegative finite-valued random variable under P_{θ_0} with expectation

$$E_{\theta_0}(Z_i) \;=\; \int_{\{f(x,\theta_0)>0\}} \frac{f(x,\theta_j)}{f(x,\theta_0)}\, f(x,\theta_0)\, dx \;\le\; 1$$

Also, for any (differentiable) concave function $g(z)$ like $\log(z)$,

$g(Z) \;\le\; g(EZ) + g'(EZ)\,(Z - EZ)$ **Taylor 2nd order, MVT form**

So $E_{\theta_0}(\log(Z_i)) \;\le\; \log(E_{\theta_0}(Z_i)) \;\le\; 0$

MLE Consistency, continued

$$\frac{1}{n} \log \{ \prod_{i=1}^{n} \frac{f(X_i, \theta_j)}{f(X_i, \theta_0)} \} \xrightarrow{P_{\theta_0}} \textbf{negative number, possibly } -\infty$$

for all $j = 1, \ldots, K$, and with P_{θ_0} probability approaching 1

$$L_n(\theta_0, \underline{X}) > \max_{j \geq 1} L_n(\theta_j, \underline{X})$$

This depends on **identifiability of** θ.

Also, this shows for some $c > 0$, with probability approaching 1,

$$L_n(\hat{\theta}; \underline{X}) \geq e^{cn} \max_{\theta \neq \hat{\theta}} L_n(\theta; \underline{X})$$

Generalization of MLE Consistency Argument

Suppose that $\Theta = C \subset \mathbb{R}^d$ is a compact set and an *iid* dataset $X_i,\ i = 1, \ldots, n$ follows a density $f(x, \theta_0)$ for unknown $\theta_0 \in \Theta$.

We already know that (when θ is identifiable):
for each $\theta_1 \in \Theta \backslash \{\theta_0\}$, $\quad E_{\theta_0}\{\log(f(X_1, \theta_1)/f(X_1, \theta_0))\} < 0$

We need to make this kind of inequality somewhat more uniform over neighborhoods to get a more general consistency result. *Following argument is due to Abraham Wald.*

Assume: **(*)** for all $\theta_0, \theta_1 \in \Theta$ with $\theta_1 \neq \theta_0$, there exists an open ball $B_r(\theta_1) = \{\theta :\ \|\theta - \theta_1\| < r\}$ such that

$$E_{\theta_0}\Big(\sup_{\theta \in B_r(\theta_1) \cap \Theta} \log f(X_1, \theta)\Big) < E_{\theta_0}(\log f(X_1, \theta_0))$$

Then it follows:

$$\frac{1}{n}\sum_{i=1}^{n} \{\, \log f(X_i,\theta_0) - \sup_{\theta\in B_r(\theta_1)\cap\Theta} \log f(X_i,\theta))\} \longrightarrow \textbf{pos. number}$$

But for every $\epsilon > 0$, the compact set $\{\theta \in \Theta : \ \|\theta - \theta_0\| \geq \epsilon\}$ is covered by finitely many such open balls $B_r(\theta_1)$.

So again only finitely many such convergence results as in the 1st display line of this slide are needed to show that with probability approaching 1 as $n \to \infty$,

$$\log L(\theta_0; \underline{X}) \ > \sup_{\theta\in\Theta:\ \|\theta-\theta_0\|>\epsilon} \log L(\theta; \underline{X})$$

This shows that for large n, with high probability any likelihood maximizer lies within ϵ of θ_0 **although the likelihood maximizer may not be unique** .

An example of the result with compact Θ is given (in Thm. 5.2.1) by a direct verification that the MLE [the relative-frequency estimator $\widehat{\theta}$ for the multinomial category probability-vector parameter $\theta = (p_1, \ldots, p_K)$] is **uniformly consistent**.