STAT5010 Advanced Statistical Inference **Fall 2023**

Lecture 8: Uniformly Most Powerful Tests

Lecturer: Tony Sit *Scribe: Yiling Kuang, Ziyao Su*

8.1 Stating The Problem

Setup: Let $\{P_\theta : \theta \in \Omega\}$ be a collection of probability measure on X, dominated by a o-finite measure μ. Let $P_\theta(\cdot) = \frac{dP_\theta}{d\mu}$. Let Ω_0 and Ω_1 be two disjoint subsets of Ω (i.e., $\Omega_1 = \Omega_0 \cup \Omega_1, \Omega_0 \cap \Omega_1 = \{\phi\}$). Given $X \sim P_\theta$ for some $\theta \in \Omega$. We have to decide whether $\theta \in \Omega_0$ or $\theta \in \Omega_1$.

Example: $X \in \mathbb{R}^n, X = (X_1 \cdots X_n)$ are i.i.d from normal distribution $N(\theta, 1), \Omega = \mathbb{R}, \Omega_0 = \{0\}$ and $\Omega_1 = \{1\}$.

A function $\phi : x \to \{0, 1\}$ is called a **non-randomized test function**, if

$$\begin{aligned} \phi = 1 &\Leftrightarrow \text{ Reject } H_0 \\ \phi = 0 &\Leftrightarrow \text{ Do not Reject } H_0 \end{aligned}$$

Probability of type I error: $P_\theta(\phi = 1), \theta \in \Omega_0$. Probability of type II error: $P_\theta(\phi = 0), \theta \in \Omega_1$. Power function of ϕ: $1 -$ probability of type II error $= P_\theta(\phi = 1), \theta \in \Omega_1$. Size of a test $\phi = \sup_{\theta \in \Omega_0} P_\theta(\phi = 1)$.

Let $\alpha \in (0, 1)$, a test ϕ is called **level** α if $\sup_{\theta \in \Omega_0} P_\theta(\phi = 1) \leq \alpha$

Def: A test ϕ is called uniformly most powerful level α test, if given any other level α test ψ, we have $P_\theta(\phi = 1) \geq P_\theta(\psi = 1), \forall \theta \in \Omega_1$.

Def: A function $\phi : x \to \{0, 1\}$ is called a randomized test function, or just a test function if $\phi(x) = p \in (0, 1)$.

Toss a coin with prob of heads p. If heads choose Ω_1, or otherwise choose Ω_0.
- Replace $P_\theta(\phi = 1)$ by $E_\theta(\phi)$. Consider the case where $\Omega_0 = \{\theta_0\}$ and $\Omega_1 = \{\theta_1\}$.

8.2 The Neyman–Pearson Fundamental Lemma

Theorem 3.2.1 (TSH): Let P_0 and P_1 be probability distributions possessing densities p_0 and p_1 respectively with respect to a measure μ.
(i) Existence. For testing $H : p_0$ against the alternative $K : p_1$ there exists a test ϕ and a constant k such that

$$E_0 \phi(X) = \alpha \tag{3.7}$$

and

$$\phi(x) = \begin{cases} 1 & \text{when} \quad p_1(x) > kp_0(x), \\ 0 & \text{when} \quad p_1(x) < kp_0(x). \end{cases} \tag{3.8}$$

(ii) Sufficient condition for a most powerful test. If a test satisfies (3.7) and (3.8) for some k, then it is most powerful for testing p_0 against p_1 at level α.
(iii) Necessary condition for a most powerful test. If ϕ is most powerful at level α for testing p_0 against p_1, then for some k it satisfies (3.8) a.e. μ. It also satisfies (3.7) unless there exists a test of size $<\alpha$ and with power 1 .

Proof: For $\alpha=0$ and $\alpha=1$ the theorem is easily seen to be true provided the value $k=+\infty$ is admitted in (3.8) and $0 \cdot \infty$ is interpreted as 0 . Throughout the proof we shall therefore assume $0<\alpha<1$.

(i): Let $\alpha(c)=P_0\left\{p_1(X)>c p_0(X)\right\}$. Since the probability is computed under P_0, the inequality need be considered only for the set where $p_0(x)>0$, so that $\alpha(c)$ is the probability that the random variable $p_1(X) / p_0(X)$ exceeds c. Thus $1-\alpha(c)$ is a cumulative distribution function, and $\alpha(c)$ is nonincreasing and continuous on the right, $\alpha(c-0)-\alpha(c)=P_0\left\{p_1(X) / p_0(X)=c\right\}, \alpha(-\infty)=1$, and $\alpha(\infty)=0$. Given any $0<\alpha<1$, let c_0 be such that $\alpha\left(c_0\right) \leq \alpha \leq \alpha\left(c_0-0\right)$, and consider the test ϕ defined by

$$\phi(x)=\begin{cases}1 & \text{when} \quad p_1(x)>c_0 p_0(x) \\ \frac{\alpha-\alpha\left(c_0\right)}{\alpha\left(c_0-0\right)-\alpha\left(c_0\right)} & \text{when} \quad p_1(x)=c_0 p_0(x) \\ 0 & \text{when} \quad p_1(x)<c_0 p_0(x)\end{cases}$$

Here the middle expression is meaningful unless $\alpha\left(c_0\right)=\alpha\left(c_0-0\right)$; since then $P_0\left\{p_1(X)=c_0 p_0(X)\right\}=0, \phi$ is defined a.e. The size of ϕ is

$$E_0 \phi(X)=P_0\left\{\frac{p_1(X)}{p_0(X)}>c_0\right\}+\frac{\alpha-\alpha\left(c_0\right)}{\alpha\left(c_0-0\right)-\alpha\left(c_0\right)} P_0\left\{\frac{p_1(X)}{p_0(X)}=c_0\right\}=\alpha,$$

so that c_0 can be taken as the k of the theorem.

(ii): Suppose that ϕ is a test satisfying (3.7) and (3.8) and that ϕ^* is any other test with $E_0 \phi^*(X) \leq \alpha$. Denote by $S^{+}$and $S^{-}$the sets in the sample space where $\phi(x)-\phi^*(x)>0$ and <0 respectively. If x is in $S^{+}, \phi(x)$ must be >0 and $p_1(x) \geq k p_0(x)$. In the same way $p_1(x) \leq k p_0(x)$ for all x in S^{-}, and hence

$$\int\left(\phi-\phi^*\right)\left(p_1-k p_0\right) d \mu=\int_{S^{+} \cup S^{-}}\left(\phi-\phi^*\right)\left(p_1-k p_0\right) d \mu \geq 0 .$$

The difference in power between ϕ and $\phi *$ therefore satisfies

$$\int\left(\phi-\phi^*\right) p_1 d \mu \geq k \int\left(\phi-\phi^*\right) p_0 d \mu \geq 0$$

as was to be proved.

(iii): Let ϕ^* be most powerful at level α for testing p_0 against p_1, and let ϕ satisfy (3.7) and (3.8). Let S be the intersection of the set $S^{+} \cup S^{-}$, on which ϕ and ϕ^* differ, with the set $\left\{x: p_1(x) \neq k p_0(x)\right\}$, and suppose that $\mu(S)>0$. Since $\left(\phi-\phi^*\right)\left(p_1-k p_0\right)$ is positive on S, it follows from Problem 2.4 that

$$\int_{S^{+} \cup S^{-}}\left(\phi-\phi^*\right)\left(p_1-k p_0\right) d \mu=\int_S\left(\phi-\phi^*\right)\left(p_1-k p_0\right) d \mu>0$$

and hence that ϕ is more powerful against p_1 than ϕ^*. This is a contradiction, and therefore $\mu(S)=0$, as was to be proved.

If ϕ^* were of size $<\alpha$ and power <1, it would be possible to include in the rejection region additional points or portions of points and thereby to increase the power until either the power is 1 or the size is α. Thus either $E_0 \phi^*(X)=\alpha$ or $E_1 \phi^*(X)=1$.

Example: Let $X_1 \dots X_n \overset{\text{ind}}{\sim} N(\theta, 1)$. Test $H_0 : \theta = 0$ us $H_1 : \theta = 1$ at level α.

$$\frac{P_{\theta=1}(X)}{P_{\theta=0}(X)} = \frac{(\frac{1}{\sqrt{2\pi}})^n exp\{-\frac{1}{2}\sum_{i=1}^n (x_i - 1)^2\}}{(\frac{1}{\sqrt{2\pi}})^n exp\{-\frac{1}{2}\sum_{i=1}^n x_i^2\}} = e^{\sum_{i=1}^n x_i - \frac{n}{2}}$$

$\Rightarrow \phi = 1$ if $\frac{P_{\theta_1}(X)}{P_{\theta_0}(X)} > K \Leftrightarrow \sum_{i=1}^n x_i - \frac{n}{2} > \log K \Leftrightarrow \sum_{i=1}^n x_i > \log K + \frac{n}{2}$

$\Rightarrow \phi(X) = \begin{cases} 1 & \text{if } \sum_{i=1}^n x_i > k' \\ 0 & \text{if } \sum_{i=1}^n x_i < k' \end{cases}$

where $\alpha = E_{\theta=0}\phi(x) = P_{\theta=0}\left(\sum_{i=1}^n X_i > k'\right)$

Example: Suppose X has a binomial distribution with success probability θ and $n = 2$ trials. If we are interested in testing $H_0 : \theta = 1/2$ versus $H_1 : \theta = 2/3$, then

$$L(X) = \frac{p_1(X)}{p_0(X)} = \frac{\begin{pmatrix} 2 \\ X \end{pmatrix} (2/3)^X (1/3)^{2-X}}{\begin{pmatrix} 2 \\ X \end{pmatrix} (1/2)^X (1/2)^{2-X}} = \frac{2^X \times 4}{9}.$$

Suppose the desired significance level is $\alpha = 50\%$. Let $\phi(x_1, x_2) = \begin{cases} 1 & \text{if } (x_1, x_2) = (1,1) \\ 0 & \text{if } (x_1, x_2) = (0,0) \end{cases}$, randomised when we observe $(0,1)$ and $(1,0)$, such that $E_{\theta_0}(\phi(X_1, X_2)) = \frac{1}{2}$.

Corrollary 3.2.1 (TSH): Let $\beta = \beta(\theta_1)$ denote the power of the Most Powerful Test for testing $H_0 : \theta = \theta_0$ and $H_1 : \theta = \theta_1$ at level $\alpha \in (0,1)$. Then $\beta \geq \alpha$. Furthermore, $\beta > \alpha$ unless $P_{\theta_1} = P_{\theta_0}$

Proof: Let ϕ be the Most Powerful Test from (i) of Neyman-Pearson lemma

$$\text{Let } \psi(x) \equiv \alpha \Rightarrow \beta = E_{\theta_1}(\phi(x)) \geq E_{\theta_1}(\psi(x)) = \alpha$$

suppose $\alpha = \beta$, then $\psi(x)$ is a Most Powerful test

$$\Rightarrow P_{\theta_1}(x) = kP_{\theta_0}(x) \text{ as } \mu_1 \Rightarrow k = 1$$
$$\Leftrightarrow P_{\theta_0} = P_{\theta_1}$$

8.2.1 Floyd-Warshall Algorithm: Dynamic Programming

Label the vertices $1, 2, \dots, n$. Define $d^{(k)}(i,j)$ to be the length of a shortest path from i to j, using intermediate vertices from $\{1, 2, \dots, k\}$ only. Obviously, $d^{(n)}(i,j)$ is the full problem.

...

Theorem 8.1 (Weak Law of Large Numbers) *Let $\boldsymbol{X} = (X_1, \dots, X_n)^\top$ be a sequence of mutually independent and identically distributed random variables, each of which has a finite mean $E(X_i) = \mu \leq \infty, i = 1, \dots, n$. Let S_n be the linear sumer of the n random variables; that is*

$$S_n = X_1 + \ldots + X_n.$$

Then for any $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) \to 0, \tag{8.1}$$

as $n \to \infty$.

8.3 Transitive Closure

Our goal is to achieve running time $O(M(n) \log n)$ for APSP where $M(n)$ is the time for $n \times n$ matrix multiplication. Let's see if we can achieve this for a simpler but related problem, namely *Transitive Closure*:

...

References

[AGM97] N. Alon, Z. Galil and O. Margalit, On the Exponent of the All Pairs Shortest Path Problem, *Journal of Computer and System Sciences* **54** (1997), pp. 255–262.

[F76] M. L. Fredman, New Bounds on the Complexity of the Shortest Path Problem, *SIAM Journal on Computing* **5** (1976), pp. 83-89.