STAT 5010: Advanced Statistical Inference

Lecturer: Tony Sit
Scribe: Yuhan HU; Yunjie Liang

Lecture 11

11 Testing of Statistical Hypothesis

Lemma 1 *Let P_0 and P_1 be probability distributions possessing densities p_0 and p_1 respectively with respect to a measure μ.*

1. *Existence. For testing $H_0 : p_0$ against the alternative $H_1 : p_1$ there exists a test $\phi(X)$ and a constant k such that*

   (a) $E_{p_0}(\phi(X)) = \alpha$

   (b) $\phi(x) = \begin{cases} 1 & p_1(x) > kp_0(x)\ [rejection] \\ 0 & p_1(x) < kp_0(x)\ [acceptance] \end{cases}$ *, such a test is called a likelihood ratio test.*

2. *Sufficiency: If a test satisfies (a) and (b) for some constant k, it is most powerful for testing $H_0 : p_0$ against $H_1 : p_1$ at level α.*

3. *Necessity: If a test ϕ is MP at level α, then it satisfies (b) for some k, and it also satisfies (a) unless there exists a test of size $< \alpha$ with power 1.*

Proof 1 *Let $\gamma(x) = \frac{p_1(x)}{p_0(x)}$ be the likelihood ratio. Denote the cumulative distribution function of $r(x)$ under H_0 as F_0.*

1. *Existence. Let $\alpha(c) = P_0(\gamma(x) > c) = 1 - F_0(c)$ for $c \in R$. Then $\alpha(c)$ is a non-increasing, right-continuous function of c $[\alpha(c) = lim_{\epsilon\to 0}\alpha(c+\epsilon)]$. Observe also that $\alpha(c)$ is not necessarily left-continuous at every value of c, but the left-hand limits exist.*
   There exists a value c_0 such that $\alpha(c_0) \le \alpha \le \alpha(c_0^-)$. Note that $F(c_0) \ge 1-\alpha \ge F(c_0^-)$, ie. c_0 is the $1-\alpha$ quantile of r(x).
   we define our test function to be:

$$\phi(x) = \begin{cases} 1 & if\ \gamma(x) > c_0, \\ \gamma & if\ \gamma(x) = c_0, \\ 0 & if\ \gamma(x) < c_0, \end{cases}$$

   for some constant γ. The test function always rejects the null if the $\gamma(x) > c_0$, and does not reject the null if $\gamma(x) < c_0$.
   The size of the test ϕ is given by

$$\begin{aligned} E_0(\phi(X)) &= P_0(\gamma(X) > c_0) + \gamma P_0(\gamma(X) = c_0) \\ &= \alpha(c_0) + \gamma\{\alpha(c_0^-) - \alpha(c_0)\} \end{aligned}$$

   If $\alpha(c_0) = \alpha(c_0^-)$ [continuous case], then $\alpha(c_0) = \alpha$ and we automatically have $E_0(\phi(X)) = \alpha$ for any choice of γ. Otherwise, we can set

$$\gamma = \frac{\alpha - \alpha(c_0)}{\alpha(c_0^-) - \alpha(c_0)}$$

   which gives also $E_0(\phi(X)) = \alpha$.

2. *Sufficiency: Let ϕ satisfies (a) and (b) in part(1), and let ϕ' be any other level α test, which satisfies*

$$E_0(\phi'(X)) = \int \phi'(x)p_0(x)d\mu(x) \leq \alpha$$

We need to bound the power difference $E_1(\phi(X)) - E_1(\phi'(X))$ from below by the size difference $E_0(\phi(X)) - E_0(\phi'(X))$ up to a constant multiple.
we claim the following inequality holds:

$$\int \{\phi(x) - \phi'(x)\}\{p_1(x) - kp_0(x)\}d\mu(x) \geq 0$$

To see this, we consider the following three cases:

   (a) *if $p_1(x) > kp_0(x)$ $[\equiv \gamma(x) > k]$, then $\phi(x) = 1$. Since $\phi'(x) \leq 1$, the the integrand is non-negative.*
   (b) *$p_1(x) < kp_0(x)$, ...*
   (c) *$p_1(x) = kp_0(x)$, ...*

It implies that

$$\int \{\phi(x) - \phi'(x)\}p_1(x)d\mu(x) \geq k \int \{\phi(x) - \phi'(x)\}p_0(x)d\mu(x) \geq 0$$

meaning that $E_1(\phi(x)) > E_1(\phi'(x))$, ie. ϕ is most powerful at level α.

3. *Necessity: Suppose ϕ^* is most powerful at level α, and let ϕ be a likelihood ratio test satisfying (a) and (b), we want to show that $\phi^*(x) = \phi(x)$ except possibly cases where $\frac{p_1(x)}{p_0(x)} = k$ for μ-a.l. x. Define*

$$S^+ = \{x : \phi(x) > \phi^*(x)\}$$
$$S^- = \{x : \phi(x) < \phi^*(x)\}$$
$$S_0 = \{x : \phi(x) = \phi^*(x)\}$$

and $S = (S^+ \cup S^-) \cap \{x : p_1(x) \neq kp_0(x)\}$. We want to show that $\mu(x) = 0$. Suppose $\mu(x) > 0$, as in part 2, we have $(\phi(x) - \phi'(x))(p_1(x) - kp_0(x)) > 0$ on S. Therefore,

$$\begin{aligned}\int_x (\phi(x) - \phi^*(x))(p_1(x) - kp_0(x))d\mu x &= \int_{S^+ \cup S^-} (\phi(x) - \phi^*(x))(p_1(x) - kp_0(x))d\mu x \\ &= \int_S (\phi(x) - \phi^*(x))(p_1(x) - kp_0(x))d\mu x > 0\end{aligned}$$

By hypothesis, $E_0(\phi(x)) = \alpha$ and $E_0(\phi^*(x)) \leq \alpha$, so the previous inequality implies that

$$E_1(\phi(x)) - E_1(\phi^*(x)) > k\{E_0(\phi(x)) - E_0(\phi^*(x))\}$$

$\Rightarrow E_1(\phi(x)) \geq E_1(\phi^*(x))$, which contradicts the assumption that ϕ^* is most powerful. Hence $\mu(S) = 0$. It remains to show that the size of ϕ^* is α unless there exists a test of size which is strictly less than α and power 1. For this, note that if size < α and power < 1, we can add points to rejection region until either the size = α or the power is 1.

Definition 1 *For simple $H_0 : P_0$ vs $H_1 : P_1$, we call $\beta_\phi(P_1) = E_{P_1}(\phi(x))$ the power of ϕ. [prob(rejecting $H_0|H_1$]).*

Corollary 2 *(TSH 3.2.1) Suppose β is the power of a most powerful level α test of H_0:P_0 vs $H_1 : P_1$ with $\alpha \in (0, 1)$. Then $\alpha < \beta$ (unless $P_0 = P_1$).*

Proof 2 *Consider the test $\phi_0(x) \equiv \alpha$, which rejects the null with probability α. Since ϕ_0 is level α, and β is the max power, we have*

$$\beta > E_1(\phi_0(x)) = \alpha$$

Suppose $\beta = \alpha$, then $\phi_0(x) = \alpha$ is a most powerful level α test. As a result,

$$\phi_0(x) = \begin{cases} 1 & if \;\; \gamma(x) > k, \\ 0 & if \;\; \gamma(x < k, \end{cases}$$

as by NP lemma 3 for same k.
Since $\phi_0(x)$ never equals 0 or 1, it must be the case that $p_1(x) = kp_0(x)$ w.p. 1. Note that

$$\int p_1(x)d\mu(x) = k \int p_0(x)d\mu(x) = 1$$

This implies that k=1. Hence $P_0 = P_1$.

Example 1 *(One parameter exponential family) Consider $X_1, ..., X_n \overset{iid}{\sim} p_\theta(x) \propto h(x)\exp(\theta T(x))$. We are interested in the testing*

$$H_0 : \theta = \theta_0 \; vs \; H_1 : \theta = \theta_1$$

The likelihood ratio is given by

$$\frac{\prod_{i=1}^n p_{\theta_1}(x_i)}{\prod_{i=1}^n p_{\theta_0}(x_i)} \propto exp((\theta_1 - \theta_0)\sum_{i=1}^n T(x_i))$$

Assuming that $\theta_1 > \theta_0$, we shall reject H_0 for large $\sum_{i=1}^n T(x_i)$. In other words, an MP test is of the form:

$$\phi(x) = \begin{cases} 1 & if \;\; \sum_{i=1}^n T(x_i) > k, \\ \gamma & if \;\; \sum_{i=1}^n T(x_i) = k, \\ 0 & if \;\; \sum_{i=1}^n T(x_i) < k, \end{cases}$$

Of course, the quantities k and γ are chosen to satisfy

$$\alpha = E_{\theta_0}\phi(x) = P_{\theta_0}(\sum_{i=1}^n T(x_i) > k) + \gamma P_{\theta_0}(\sum_{i=1}^n T(x_i) = k)$$

Note also that $\sum_{i=1}^n T(x_i)$ has no θ dependence and that k and γ do not depend on θ_1 (assuming that $\theta_1 > \theta_0$ only). This means that ϕ is in fact uniformly MP for testing:

$$H_0 : \theta = \theta_0 \; vs \; H_1 : \theta > \theta_0$$

Monotone Likelihood Ratios (MLR) and UMP one-sided Tests.

Definition 3 *We say that the family of densities $\{p_\theta : \theta \in R\}$ has monitone likelihood ratio in T(x) if*

1. *$\theta \neq \theta'$ implies $p_\theta \neq p_{\theta'}$ (identifiability)*
2. *$\theta < \theta'$ implies $p_{\theta'}(x)/p_\theta(x)$ is a non-decreasing function of T(x) (Monotonicity)*

Example 2 *(Double exponential) Let X ~ Double Exponential (θ) with density $p_\theta(x) = \frac{1}{2}e^{-|x-\theta|}$. To check the second cond, fix any $\theta' > \theta$ and consider*

$$\frac{p_{\theta'}(x)}{p_\theta(x)} = e^{|x-\theta|-|x-\theta'|}$$

observe that

$$|x-\theta| - |x-\theta'| = \begin{cases} \theta - \theta' & if \;\; x < \theta \\ 2x - \theta - \theta' & if \;\; \theta \le x \le \theta' \\ \theta' - \theta & if \;\; x > \theta', \end{cases}$$

Which is non-decreasing in x. Therefore the family has MLR in T(x)=x.

Example 3 *(Cauchy location model). Let X have density $p_\theta = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2}$. We consider the likelihood ratio if $\frac{p_\theta(x)}{p_{\theta'}(x)} = \frac{1+x^2}{1+(x-\theta)^2} \to 1$ as $x \to \infty$ or $x \to -\infty$ for $\theta > 0$. But $p_\theta(0)/p_{\theta'}(0) = \frac{1}{1+\theta^2} < 1$. This ratio must increase at some value of x and decrease at other locations. Hence this family does not satisfy the MLR property.*

Theorem 4 *(TSH 3.4.1) Suppose $X \sim p_\theta(x)$ has MLR is T(x) and we test $H_0 : \theta \le \theta_0$ vs $H_1 : \theta > \theta_0$. Then,*

1. *there exists a UMP test at level α of the form*

$$\phi(x) = \begin{cases} 1 & if \;\; T(x) > k \\ \gamma & if \;\; T(x) = k \\ 0 & if \;\; T(x) < k, \end{cases}$$

   where k and γ are determined by $E_{\theta_0}\phi(x) = \alpha_j$.

2. *the power function $\beta(\theta) = E_\theta\phi(X)$ is strictly increasing when $0 < \beta(\theta) < 1$.*

Outline 5 *To show ϕ is UMP at level α for testing $H_0 : \theta \le \theta_0$ vs $H_1 : \theta > \theta_0$, we have to show that the constraint for $\theta < \theta_0$.*

Optinal Test for composite Nulls. Consider the case with a simple alternative:

$$H_0 : X \sim f_0, \theta \in \Omega_0$$

$$H_1 : X \sim g(unknown), [simple]$$

We impose a prior distribution Λ on Ω_0. So we consider the new hypothesis:

$$H_\Lambda : X \sim h_\Lambda(x) = \int_{\Omega_0} f_0(x) d\Lambda(\theta)$$

, where $h_\Lambda(x)$ is the marginal distribution of X induced by Λ. We shall test H_Λ vs H_1.
Let β_Λ be the power of the MP level-α test Φ_Λ for testing H_Λ vs. $H_1(g)$.

Definition 6 *The prior Λ is a least favourable distribution if $\beta_\Lambda \le \beta_{\Lambda'}$ for any prior Λ'.*

Theorem 7 *(TSH 3.8.1) Suppose Φ_Λ is a MP level-α test for testing H_Λ against g. If ϕ_Λ is level-α for the original hypothesis H_0 (i.e $E_{\theta_0}\Phi_\Lambda(x) \le \alpha \forall \theta \in \Omega_0$), then*

1. *The test Φ_Λ is MP for the original: $H_0 : \theta \in \Omega_0$ vs $H_1 : g$*

2. *The distribution Λ is least favourable.*

Proof 3 *:*

1. *Let Φ^* be any other level-α test of $H_0 : \theta \in \Omega_0$ vs g. Then Φ^* is also a level-α test for H_Λ vs g because*

$$E_\theta(\Phi^*(X)) = \int \Phi^*(x) f_\theta(x) d\mu(x) \leq \alpha \quad \forall \theta \in \Omega_0$$

which implies that

$$\int \Phi^*(x) h_\Lambda(x) d\mu(x) = \int \int \Phi^*(x) f_\theta(x) d\mu(x) d\Lambda(\theta) \leq \int \alpha d\Lambda(\theta) = \alpha$$

Since Φ_Λ is MP for H_Λ vs g, we have

$$\int \Phi * (x) g(x) d\mu(x) \leq \int \Phi_\Lambda(x) g(x) d\mu(x),$$

Hence Φ_Λ is a MP test for H_0 vs g because Φ_Λ is also level α.

2. *Let Λ' be any distribution on Ω_0. Since $E_\theta(\Phi_\Lambda(x) < \alpha \quad \forall \theta \in \Omega_0$, we know that Φ_Λ must be level α for $H_{\Lambda'}$ vs g. Thus $\beta_\Lambda \leq \beta_{\Lambda'}$, so Λ is the least favourable distrubution.*

Example 4 *Let $X_1, ..., X_n \overset{iid}{\sim} N(\theta, \sigma^2)$ with both σ^2 and θ unknown. We consider teting $H_0 : \sigma \leq \sigma_0$ against $H_1 : \sigma > \sigma_0$. Our goal is to find an UMP test.*

1. *Fix a simple alternative(θ_1, σ_1) for some arbitrary θ_1 and $\sigma_1 > \sigma_0$*

2. *Choose a prior Λ to collapse one null hypothesis. The least favourable prior should make the alt. Hypothesis hand to distinguish for the null. Hence a rule of thumb: to concentrate Λ on the boundary between H_1 and H_0 (i.e. the $\{\sigma = \sigma_0\}$). In this case, we assign Λ to be a prob distribution on $\theta \in R$ and a fixed $\sigma = \sigma_0$.*
   Given away test function $\Phi(x)$ and a sufficient statistic T, there exists a test function η that has the same power as Φ but depends only on X through T:

$$\eta(T(x)) = E(\Phi(x)|T(x)).$$

   We restrict our attention to (Y, u) where $Y = \bar{X}$ and $u = \sum_{i=1}^n (x_i - \bar{x})^2$. We know that $Y \sim N(0, \frac{\sigma^2}{n})$, $u \sim \sigma^2 \chi^2_{n-1}$ and Y $\perp\!\!\!\perp$ u by Basu Theorem.
   Thus for Λ supported on $\sigma = \sigma_0$, we obtain the joint density of (Y,u) under H_Λ as

$$C_0 u^{\frac{n-3}{2}} e^{-\frac{u}{2\sigma_0^2}} \int e^{-\frac{n}{2\sigma_0^2}(y-\theta)^2} d\Lambda(\theta)$$

   and the joint density under alternative (θ_1, σ_1) as

$$C_1 u^{\frac{n-3}{2}} e^{-\frac{u}{2\sigma_1^2}} \int e^{-\frac{n}{2\sigma_1^2}(y-\theta)^2} d\Lambda(\theta_1)$$

   We can see that the choice of Λ only affects the distribution of Y. To achieve the minimal max, power against the alternative, we need to choose Λ that the two distribution because as close as possible.

Under the alternative hypothesis, $Y \sim N(\theta_1, \frac{\sigma_1^2}{n})$ under H_0, the distribution of Y is a convolution from, i.e. $Y = Z + \Theta$, where $Z \sim N(0, \frac{\sigma_1^2}{n})$, $\Theta \sim \Lambda$, with $Z \perp\!\!\!\perp \Theta$, Hence if we choose $\Theta \sim N(\theta_1, \frac{\sigma_1^2 - \sigma_0^2}{n})$, Y will have the same dist, under the null and the alternative which is $N(\theta_1, \frac{\sigma_1^2}{n})$. Under this choice of prior, the LRT rejects for large value of $\exp\{-\frac{u}{2\sigma_1^2} + \frac{u}{2\sigma_0^2}\}$ (hence large value of u). So the MP test rejects H_Λ if $\sum_{i=1}^n (x_i - \bar{x})^2 > \sigma_0^2 C_{n-1,1-\alpha}$

3. *We need to check if the MP test is for the composite null. For any (θ, σ) with $\sigma < \sigma_0$, the prob of rejection is*

$$P_{\theta,\sigma}\left(\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2}\right) = P\left(\chi^2_{n-1} > \frac{\sigma_0^2}{\sigma^2} C_{n-1,1-\alpha}\right) \leq \alpha$$

with equality holds $\sigma = \sigma_0$. Hence our test is MP for the testing original null H_0 vs $N(\theta_1, \sigma_1)$.

4. *The test does not depend on (θ_1, σ_1). Hence the test above is NMP for testing the original null vs the composite alternative.*