STATS 300A: Theory of Statistics	Fall 2015
Lecture 13 — November 05	
Lecturer: Lester Mackey	*Scribe: Martin Jinye Zhang, Can Wang*

Warning: *These notes may contain factual and/or typographic errors.*

13.1 Neyman-Pearson Lemma

Recall that a hypothesis testing problem consists of the **data** $X \sim \mathbb{P}_\theta \in \mathcal{P}$, a **null hypothesis** $H_0 : \theta \in \Omega_0$, an **alternative hypothesis** $H_1 : \theta \in \Omega_1$, and the set of candidate **test functions** $\phi(x)$ representing the probability of rejecting the null hypothesis given the data x.

Given a significance level α, our optimality goal is to maximize the power $\mathbb{E}_{\theta_1}\phi(x)$, for every $\theta_1 \in \Omega_1$, subject to the size constraint, (i.e. size $\leq$ level),

$$\mathbb{E}_{\theta_0}\phi(x) \leq \alpha \quad \text{for every } \theta_0 \in \Omega_0.$$

In many hypothesis testing problems, the goal of simultaneously maximizing the power under every alternative is unachievable. However, we saw last time that the goal can be achieved when both the null and alternative hypotheses are simple, via the **Neyman-Pearson Lemma**.

Theorem 1 (Neyman-Pearson Lemma (TSH 3.2.1))**.**

(i) *Existence.* For testing simple $H_0 : p_0$ against simple $H_1 : p_1$, there exist a test function ϕ and a constant k such that

  (a) $\mathbb{E}_{p_0}\phi(X) = \alpha$, and

  (b) ϕ has the form

$$\phi(x) = \begin{cases} 1 & \text{if } \frac{p_1(x)}{p_0(x)} > k \\ 0 & \text{if } \frac{p_1(x)}{p_0(x)} < k. \end{cases}$$

(ii) *Sufficiency.* If ϕ satisfies (a) and (b) for some constant k, then ϕ is most powerful at level α.

(iii) *Necessity.* If a test ϕ^* is most powerful at level α, then it satisfies (b) for some k, and it also satisfies (a) unless there exists a test of size strictly less than α with power 1.

Note that part (a) of (i) states that the size of ϕ is exactly equal to the significance level, and part (b) states that ϕ takes the form of a likelihood ratio test.

Proof. Let $r(x) = \dfrac{p_1(x)}{p_0(x)}$ be the likelihood ratio and denote the cumulative distribution function of $r(X)$ under H_0 by F_0.

(i) *Existence.* Let $\alpha(c) = \mathbb{P}_0(r(X) > c) = 1 - F_0(c)$. Then $\alpha(c)$ is a non-increasing, right-continuous (i.e., $\alpha(c) = \lim_{\epsilon \searrow 0} \alpha(c + \epsilon)$) function of c. Note that $\alpha(c)$ is not necessarily left-continuous at every value of c, but the left-hand limits exist; denote them by $\alpha(c^-) = \lim_{\epsilon \searrow 0} \alpha(c - \epsilon)$.

The above properties of α imply that there exists a value c_0 such that $\alpha(c_0) \leq \alpha \leq \alpha(c_0^-)$ (Figure 13.1). Note that $F_0(c_0) \geq 1 - \alpha \geq F_0(c_0^-)$, i.e., c_0 is the $(1-\alpha)$ quantile of $r(X)$.

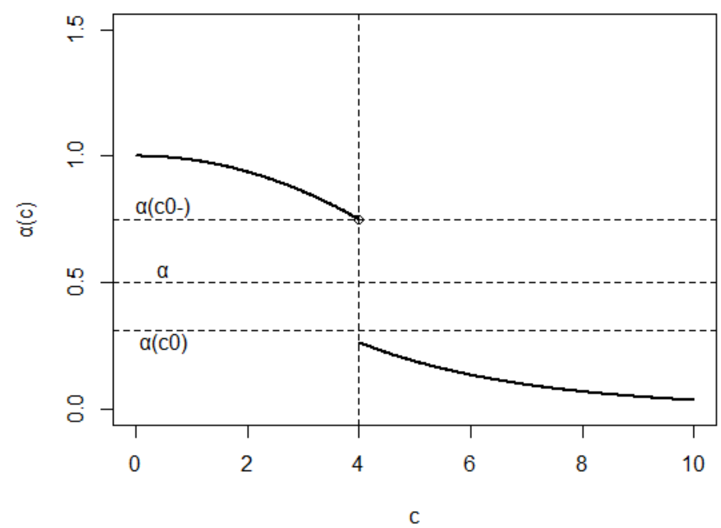

Figure 13.1. The existence of c_0 satisfying $\alpha(c_0) \leq \alpha \leq \alpha(c_0^-)$

We define our test function to be

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

for some constant γ. We note here that ϕ only depends on x through $r(x)$. ϕ always rejects if the likelihood ratio strictly exceeds the threshold c_0, never rejects if the likelihood ratio falls strictly below c_0, and rejects with some probability γ whenever the likelihood ratio equals c_0. By construction, if we take $k = c_0$, then ϕ satisfies condition (b) of part (i). We now choose γ so that ϕ also satisfies part (a).

The size of ϕ is given by

$$\begin{aligned} \mathbb{E}_0\phi(X) &= \mathbb{P}_0(r(X) > c_0) + \gamma \mathbb{P}_0(r(X) = c_0) \\ &= \alpha(c_0) + \gamma[\alpha(c_0^-) - \alpha(c_0)]. \end{aligned}$$

If $\alpha(c_0^-) = \alpha(c_0)$, then $\alpha(c_0) = \alpha$ and we automatically have $\mathbb{E}_0\phi(X) = \alpha$ for any choice of γ. Otherwise, set

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

which gives $\mathbb{E}_0\phi(X) = \alpha$.

(ii) *Sufficiency.* Let ϕ satisfy (a) and (b) of part (i), and let ϕ' be any other level α test, so that

$$\mathbb{E}_0[\phi'(X)] = \int \phi'(x)p_0(x)d\mu(x) \le \alpha.$$

In order to show ϕ is most powerful, we bound the power difference $\mathbb{E}_1\phi(X) - \mathbb{E}_1\phi'(X)$ from below by the size difference $\mathbb{E}_0\phi(X) - \mathbb{E}_0\phi'(X)$ up to a constant multiple, whence the result follows since $\mathbb{E}_0\phi(X) = \alpha$. To do this, we claim that the following inequality holds:

$$\int (\phi(x) - \phi'(x))(p_1(x) - kp_0(x))d\mu(x) \ge 0.$$

To see this, we consider three cases:

(a) If $p_1(x) > kp_0(x)$, then $\phi(x) = 1$ (by construction). Since $\phi'(x) \le 1$, the integrand is nonnegative.

(b) If $p_1(x) < kp_0(x)$, then $\phi(x) = 0$ (by construction). Since $\phi'(x) \ge 0$, the integrand is nonnegative.

(c) If $p_1(x) = kp_0(x)$, then the integrand is zero.

By exhaustion, the inequality holds. Rearranging terms, we have

$$\begin{aligned} \int (\phi(x) - \phi'(x))p_1(x)d\mu(x) \quad &\ge \quad k\int (\phi(x) - \phi'(x))p_0(x)d\mu(x) \\ &= \quad k[\underbrace{\mathbb{E}_0\phi(X)}_{\alpha} - \underbrace{\mathbb{E}_0\phi'(X)}_{\le\alpha}] \ge 0. \end{aligned}$$

We conclude that $\mathbb{E}_1\phi(X) \ge \mathbb{E}_1\phi'(X)$; i.e., ϕ is most powerful at level α.

(iii) *Necessity.* Suppose ϕ^* is most powerful at level α, and let ϕ be a likelihood ratio test satisfying (a) and (b) of part (i). We must show that $\phi^*(x) = \phi(x)$ except possibly where $p_1(x)/p_0(x) = k$, for μ-a.e. x. Define the sets

$$\begin{aligned} S^+ &= \{x : \phi(x) > \phi^*(x)\}, \\ S^- &= \{x : \phi(x) < \phi^*(x)\}, \\ S_0 &= \{x : \phi(x) = \phi^*(x)\}, \end{aligned}$$

and

$$S = (S^+ \cup S^-) \cap \{x : p_1(x) \ne kp_0(x)\}.$$

We must show that $\mu(S) = 0$. Suppose not, so that $\mu(S) > 0$. As in part (ii), we have $(\phi - \phi^*)(p_1 - kp_0) > 0$ on S. Therefore,

$$\int_{\mathcal{X}} (\phi - \phi^*)(p_1 - kp_0) d\mu(x) = \int_{S^+ \cup S^-} (\phi - \phi^*)(p_1 - kp_0) d\mu(x) = \int_S (\phi - \phi^*)(p_1 - kp_0) d\mu(x) > 0,$$

where the two equalities above hold because the integrand on their difference sets is equal to 0. By hypothesis, $\mathbb{E}_0 \phi(X) = \alpha$ and $\mathbb{E}_0 \phi^*(X) \leq \alpha$, so the previous inequality implies

$$\mathbb{E}_1 \phi(X) - \mathbb{E}_1 \phi^*(X) > k[\mathbb{E}_0 \phi(X) - \mathbb{E}_0 \phi^*(X)] \geq 0.$$

That is, $\mathbb{E}_1 \phi(X) > \mathbb{E}_1 \phi^*(X)$, 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 strictly less than α and power 1. For this, note that if size $< \alpha$ and power < 1, we can add points to rejection region until either the size is α or the power is 1. □

Definition 1. For simple $H_0 : P_0$ vs $H_1 : P_1$, we call $\beta_\phi(P_1) = \mathbb{E}_{P_1}[\phi(x)]$ the power of ϕ, i.e. the probability of rejection under the alternative hypothesis.

Corollary 1 (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 $P_0 = P_1$).

The takeaway is that a MP test rejects more often under the alternative hypothesis than under the null hypothesis, which is an appealing property for a test to have.

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

$$\beta \geq \mathbb{E}_{P_1} \phi_0(X) = \alpha.$$

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

$$\phi_0(x) = \begin{cases} 1 & \text{if } p_1(x)/p_0(x) > k, \\ 0 & \text{if } p_1(x)/p_0(x) < k, \end{cases} \quad \text{a.s. by N-P (iii), for some } k.$$

Since $\phi_0(x)$ never equals 0 or 1, it must be the case that $p_1(X) = k \cdot p_0(X)$ with probability 1. Note that

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

This implies that $k = 1$ and hence $P_0 = P_1$. □

13.2 Exponential Families and UMP One-sided Tests

In certain cases, we can boost MP tests for a simple alternative up to UMP tests for a composite alternative. We gave an example for the normal distribution last time; here we provide a more general example.

Example 1 (One parameter exponential family)**.** Consider the case where $X_1, \dots, X_n \overset{i.i.d.}{\sim} p_\theta(x) \propto h(x) \exp(\theta T(x))$, and we are interested in testing

$$H_0 : \theta = \theta_0 \quad vs. \quad H_1 : \theta = \theta_1.$$

We want an MP test at level α. The likelihood ratio is

$$\frac{\prod_i p_{\theta_1}(x_i)}{\prod_i p_{\theta_0}(x_i)} \propto \exp\left((\theta_1 - \theta_0)\sum_i T(x_i)\right).$$

Since the exponential is just a monotone transformation, an MP test will reject for large $(\theta_1 - \theta_0)\sum_i T(x_i)$. Assuming $\theta_1 > \theta_0$, we will reject for large $\sum_i T(x_i)$. That is, an MP test has the form

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

where k, γ are chosen to satisfy the size constraint

$$\alpha = \mathbb{E}_{\theta_0}\phi(X) = \mathbb{P}_{\theta_0}\left[\sum_i T(X_i) > k\right] + \gamma \mathbb{P}_{\theta_0}\left[\sum_i T(X_i) = k\right].$$

Note that $\sum_i T(x_i)$ has no explicit θ dependence and that k, γ do not depend on θ_1 (assuming $\theta_1 > \theta_0$). This means ϕ is in fact UMP for testing

$$H_0 : \theta = \theta_0 \quad vs. \quad H_1 : \theta > \theta_0.$$

Here, H_1 is an example of a *one-sided alternative*, which arises when the parameter values of interest lie on only one side of the real-valued parameter θ_0.

13.3 Monotone Likelihood Ratios and UMP One-sided Tests

In the above example, we were able to extend our MP test for a simple hypothesis to a UMP test for a one-sided hypothesis. This phenomenon is not unique to exponential families. We can get the same behavior whenever the models have a so-called *monotone likelihood ratio.*

Definition 2 (Families with monotone likelihood ratio (MLR))**.** We say that the family of densities $\{p_\theta : \theta \in \mathbb{R}\}$ has **monotone likelihood ratio** in $T(x)$ if

(i) $\theta \neq \theta'$ implies $p_\theta \neq p_{\theta'}$ (identifiability),

(ii) $\theta < \theta'$ implies $p_{\theta'}(x)/p_\theta(x)$ is a nondecreasing function of $T(x)$ (monotonicity).

The one-parameter exponential family of Example 1 has MLR in the sufficient statistic $T(x) = \sum_i T(x_i)$. Here is another example.

Example 2 (Double exponential)**.** Let $X \sim \text{DoubleExponential}(\theta)$, with density $p_\theta(x) = \frac{1}{2}\exp(-|x-\theta|)$. It is easy to see that the model is identifiable, so we need only check the second condition. Fix any $\theta' > \theta$ and consider the likelihood ratio

$$\frac{p_{\theta'}(x)}{p_\theta(x)} = \exp\left(|x-\theta| - |x-\theta'|\right).$$

Note that

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

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

Finally, we give an example of a model that does not exhibit MLR in $T(x) = x$.

Example 3 (Cauchy location model)**.** Let X have density $p_\theta(x) = \dfrac{1}{\pi} \cdot \dfrac{1}{1+(x-\theta)^2}$. We find two points for which the MLR condition fails. For any fixed $\theta > 0$,

$$\frac{p_\theta(x)}{p_0(x)} = \frac{1+x^2}{1+(x-\theta)^2} \to 1 \qquad \text{as } x \to \infty \text{ or } x \to -\infty,$$

but $p_\theta(0)/p_0(0) = 1/(1+\theta^2)$, which is strictly less than 1. Thus the ratio must increase at some values of x and decrease at others. In particular, it is not monotone in x. Here we have shown that the likelihood ratio in $T(x) = x$ is not MLR.

When we have a MLR, we can boost our simple MP tests to UMP tests for certain composite hypotheses.

Theorem 2 (TSH 3.4.1)**.** Suppose $X \sim p_\theta(x)$ has MLR in $T(x)$ and we test $H_0 : \theta \le \theta_0$ vs. $H_1 : \theta > \theta_0$. Then

(i) There exists a UMP test at level α of the form

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

where k, γ are determined by the condition $\mathbb{E}_{\theta_0}\phi(X) = \alpha$.

(ii) The power function $\beta(\theta) = \mathbb{E}_\theta \phi(X)$ is strictly increasing when $0 < \beta(\theta) < 1$.

Note: The proof of the theorem relies on Corollary 1 to the NP lemma. We have seen (in Example 1) why ϕ is UMP at level α for $H_0' : \theta = \theta_0$ vs. $H_1' : \theta > \theta_0$. To show that 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 size constraint is satisfied for $\theta < \theta_0$. Part (ii) implies the power function is strictly increasing, so if ϕ achieves level α at θ_0, it rejects less when $\theta < \theta_0$. Thus, ϕ is also level α for $H_0 : \theta \le \theta_0$.