LECTURE 9: UNIFORMLY MOST POWERFUL TESTS

Michal Kolesár*

September 30, 2024

REFERENCE CB, Chapter 8.2.1 and 8.3.2, HMC, Chapter 8.1–2, H22, Chapter 13.10–12

1. NEYMAN-PEARSON LEMMA

Consider testing the null $H_0\colon \theta \in \Theta_0$ against the alternative $H_1\colon \theta \in \Theta_1$ based on data $X \sim F(\cdot \mid \theta)$. Recall that ideally, we'd like to find a test ϕ that achieves maximum power $E_\theta[\phi(X)]$ for $\theta \in \Theta_1$ subject to the level constraint

$$E_\theta \phi(X) = P_\theta(X \in S_1) \leq \alpha \quad \text{for all } \theta \in \Theta_0.$$

In general this is a hard problem because it involves optimizing over the space of functions $\{\phi\colon \phi(x) \in \{0, 1\}, x \in \mathcal{X}\}$, and we're not even sure that a uniformly most powerful test exists: it may be that for any test ϕ that achieves the maximum possible power at some $\theta \in \Theta_1$, we can find another test ϕ' with better power at some other $\theta' \in \Theta_1$.

If both the null and alternative are simple, then this problem turns out to have an elegant solution. To state it, we need to again allow for randomized decision rules by allowing for *randomized tests* $\phi\colon \mathcal{X} \to [0, 1]$. When $0 < \phi(x) < 1$, this means that we reject then null with probability $\phi(x)$. Randomized tests are seldom used in practice, but they are a useful theoretical device.

The next result is due to Neyman and Pearson (1933), published in the world's oldest scientific journal (established in 1665). Egon Pearson was Karl Pearson's son.

Lemma 1 (Neyman-Pearson, HMC Theorem 8.1.1, CB Theorem 8.3.12). Let P_0 and P_1 be distributions with densities f_0 and f_1. For testing $H_0\colon P_0$ against the alternative $H_1\colon P_1$, the test that

*Email: mkolesar@princeton.edu.

satisfies, for some $\lambda \geq 0$ and $\kappa \in [0,1]$,

$$\phi(x) = \begin{cases} 1 & \text{if } f_1(x) > \lambda f_0(x), \\ \kappa & \text{if } f_1(x) = \lambda f_0(x), \\ 0 & \text{if } f_1(x) < \lambda f_0(x). \end{cases} \tag{1}$$

is most powerful. Here κ and λ are chosen to satisfy $E_{P_0}[\phi(X)] = \alpha$.

Proof. Let $\phi'(x)$ be any other test with level α. We need to show that

$$\int_{\mathcal{X}} \phi'(x) f_1(x)\, \mathrm{d}x \leq \int_{\mathcal{X}} \phi(x) f_1(x)\, \mathrm{d}x.$$

Note that, by definition of $\phi(x)$,

$$(\phi(x) - \phi'(x))(f_1(x) - \lambda f_0(x)) \geq 0,$$

and hence, integrating wrt x,

$$\begin{aligned} 0 &\leq \int_{\mathcal{X}} (\phi(x) - \phi'(x))(f_1(x) - \lambda f_0(x))\, \mathrm{d}x \\ &= \int_{\mathcal{X}} \phi(x) f_1(x)\, \mathrm{d}x - \lambda \int_{\mathcal{X}} \phi(x) f_0(x)\, \mathrm{d}x - \int_{\mathcal{X}} \phi'(x) f_1(x)\, \mathrm{d}x + \lambda \int_{\mathcal{X}} \phi'(x) f_0(x)\, \mathrm{d}x \\ &\leq \int_{\mathcal{X}} \phi(x) f_1(x)\, \mathrm{d}x - \lambda\alpha - \int_{\mathcal{X}} \phi'(x) f_1(x)\, \mathrm{d}x + \lambda\alpha = \int_{\mathcal{X}} \phi(x) f_1(x)\, \mathrm{d}x - \int_{\mathcal{X}} \phi'(x) f_1(x)\, \mathrm{d}x, \end{aligned}$$

which proves the result. □

Remark 2. From this proof, it also follows that any other test ϕ' that is also most powerful has to satisfy (i) $\phi'(x) = 1$ if $f_1(x) > \lambda f_0(x)$, and $\phi'(0) = 0$ if $f_1(x) < \lambda f_0(x)$ for almost all x, and (ii) $E_{P_0}[\phi'(x)] = \alpha$. To see this, if ϕ' is also most powerful, then it follows from the proof that we must have $(\phi(x) - \phi'(x))(f_1(x) - \lambda f_0(x)) = 0$ for almost all x, and hence that $\phi(x) = \phi'(x)$ for almost all x such that $f_1(x) \neq \lambda f_0(x)$. The only degree if freedom is what to do if $f_1(x) = \lambda f_0(x)$.

Furthermore, it also follows that any test that has the form in eq. (1), for some κ and λ must be most powerful at level $\tilde{\alpha} = E_{P_0}[\phi(x)]$.

- The function $LR(x) = f_1(x)/f_0(x)$ is called the *likelihood ratio*. The lemma says that the most powerful test rejects for large values of the likelihood ratio (if the support of x under the null is smaller, then this is still true if we define $LR(x) = \infty$ for $f_0(x) = 0$).
- Here the randomization device just serves to make the size equal to α if it's not possible to guarantee that otherwise. It's not necessary when the likelihood ratio is continuously distributed.

Example 1. Let $X_1, \dots, X_n$ be a random sample from $\mathcal{N}(\theta, \sigma^2)$ distribution with known

σ^2. Suppose we want to test $H_0\colon \theta = \theta_0$ against $H_1\colon \theta = \theta_1$. The likelihood ratio is

$$\begin{aligned}\frac{\prod_{i=1}^n \exp(-\frac{1}{2\sigma^2}(X_i - \theta_1)^2)}{\prod_{i=1}^n \exp(-\frac{1}{2\sigma^2}(X_i - \theta_0)^2)} &= \exp\left(-\frac{1}{2\sigma^2}\left(\sum_i (X_i - \theta_1)^2 - \sum_i (X_i - \theta_0)^2\right)\right) \\ &\propto \sum_i (X_i - \theta_0)^2 - \sum_i (X_i - \theta_1)^2 \\ &\propto \overline{X}_n(\theta_1 - \theta_0)\end{aligned}$$

By the Neyman-Pearson lemma, the most powerful test therefore rejects for large values of $\overline{X}_n$ if $\theta_1 > \theta_0$, and small values of $\overline{X}_n$ if $\theta_0 < \theta_1$. Consider the first case. Then the most powerful test rejects whenever $\overline{X}_n \geq c$, where the critical value is chosen to satisfy the level constraint, $P_{\theta_0}(\overline{X} \geq c) = \alpha$, which gives $c = \theta_0 + \sigma z_{1-\alpha}/\sqrt{n}$, where $z_{1-\alpha}$ is the $1 - \alpha$ quantile of standard normal distribution. So the usual one-sided z-test is most powerful. ⊠

Example 2. Let X be a scalar with distribution $f_0(x) = \mathbb{1}\{0 \leq x \leq 1\}$ under the null, and $f_1(x) = \mathbb{1}\{0 \leq x \leq 1\}2x$ under the alternative. Here the likelihood ratio is $2x$ on the support of X, so that the test rejects for large values of X. Since X is uniform under the null, the appropriate critical value is $1 - \alpha$, which yields $\phi(X) = \mathbb{1}\{X \geq 1 - \alpha\}$. ⊠

Example 3. Suppose we wish to test the null $H_0\colon \theta = 1/2$ against the alternative $H_1\colon \theta = 3/4$, based on observing n draws of a Bernoulli, $X_i \sim \text{Bern}(\theta)$, $i = 1, 2$. Let $S = \sum_{i=1}^n X_i$ denote the sum. The joint likelihood is $\theta^S(1-\theta)^{n-S}$, so that

$$LR(X_1, X_2) = \left(\frac{3/4}{1/2}\right)^S \left(\frac{1/4}{1/2}\right)^{2-S} = (3/2)^S(1/2)^{n-S} = 3^S(1/2)^n.$$

So we reject for large values of S. To satisfy the level constraint, we will generally need to randomize. For example, if $n = 2$, and $\alpha \leq 1/4$, we reject with probability $\kappa = 4\alpha$ if $S = 2$, and don't reject otherwise. ⊠

2. COMPOSITE HYPOTHESES

The Neyman-Pearson lemma solves the problem of testing a simple null against a simple alternative. This is usually not that interesting in practice. In some special cases, the lemma can be extended to find uniformly most powerful (UMP) tests in other scenarios. Here we discuss one such scenario. First a definition:

Definition 3. A family of densities $f(x \mid \theta)$ with $\theta \in \mathbb{R}$ has a *monotone likelihood ratio* if there exists some function $T(x)$ such that for any $\theta < \theta'$, the ratio $f(x \mid \theta')/f(x \mid \theta)$ is a non-decreasing function of $T(x)$.

Theorem 4 (HMC, p. 443, CB, Theorem 8.3.17). Let $f(x \mid \theta)$, $\theta \in \mathbb{R}$ be some parametric family with monotone likelihood ratio in $T(x)$. Consider testing the null $H_0\colon \theta \leq \theta_0$ against the *alternative $H_1\colon \theta > \theta_0$ at level α. Then the test that satisfies*

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

is UMP. Here the constants κ and λ chosen to satisfy $E_{\theta_0}[\phi(X)] = \alpha$. In addition, the power of this test $\beta(\theta) = E_\theta[\phi(X)]$ is strictly increasing in θ for all θ such that $\beta(\theta) \in (0, 1)$.

Proof. Fix some simple alternative $\theta_1 > \theta_0$. By the Neyman-Pearson lemma, the UMP test of θ_0 against θ_1 rejects for large values of $f(X \mid \theta_1)/f(X \mid \theta_0)$. By the monotone likelihood ratio property, this ratio depends on X through $T(X)$, and is monotone in $T(X)$. Therefore, the test rejects for large values of $T(X)$, which does not depend on the particular alternative θ_1.

To show that the power function is increasing in θ, observe that by Remark 2, there exists a level $\tilde{\alpha}$ such that this test is also most powerful for $H_0\colon \tilde{\theta}_0$ against $H_1\colon \theta = \tilde{\theta}_1$, for any $\tilde{\theta}_0 < \tilde{\theta}_1$. It also must be the case that $\tilde{\alpha} = \beta(\tilde{\theta}_0) \leq \beta(\tilde{\theta}_1)$, since the test must be more powerful than the test $\tilde{\phi}(x) = \tilde{\alpha}$. Furthermore, the inequality must be strict, since otherwise $\tilde{\phi}(x)$ would also be most powerful, which by Remark 2 implies that $f(x, \tilde{\theta}_0) = f(x; \tilde{\theta}_1)$ for almost all x, but that is ruled out by the monotone likelihood ratio property. So it must be that the power function is strictly increasing, from which it also follows that the test controls size. □

Example 1 (continued). This distribution has monotone likelihood ratio, so the (one-sided) z-test is UMP for testing $H_0\colon \theta \leq \theta_0$ against $H_1\colon \theta > \theta_0$. ⊠

Digression. By generalizing this example, one can show that in a linear regression model $Y_i = X_i'\beta + \epsilon_i$, with $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, and σ^2 known, the one-sided z-test based on $\hat{\beta}_1$ is UMP for testing $H_0\colon \beta_1 \leq \beta_{1,0}$ against the alternative $H_1\colon \beta_1 \geq \beta_{1,0}$. *This is why we use z-tests in linear regression analysis.* ⊠

Example 3 (continued). The binomial distribution also has a monotone likelihood ratio, so tests of the hypothesis $H_0\colon \theta \leq \theta_0$ against $H_1\colon \theta > \theta_0$ reject for large values of S. ⊠

3. UNBIASEDNESS

In most cases, no UMP test exists. The simplest example is when we test against two-sided alternatives:

Example 4. Let $X_1, \ldots, X_n$ be a random sample from $\mathcal{N}(\theta, \sigma^2)$ distribution with known σ^2. Suppose we want to test $H_0\colon \theta = \theta_0$ against the alternative hypothesis, $H_1\colon \theta \neq \theta_0$. We know that the UMP test against alternatives that are bigger than θ_0 reject for large values of $\overline{X}_n$, but for alternatives smaller than θ_0, we'd want to reject for small values of $\overline{X}_n$. So no test can simultaneously be UMP against alternatives Θ_1 that contain values both bigger and smaller than θ_0. ⊠

When no UMP test exists, we have a couple of options of finding a test with good finite sample properties:

1. impose unbiasedness;

2. impose some invariance requirements;

3. use weighted average power, weighted by some function $w(\theta)$ on Θ_1 (average risk approach), or seek to maximize the minimum power under the alternative (minimax approach).

We'll only discuss the first option here.

Definition 5. A level α test ϕ is *unbiased* if its power function satisfies $\beta(\theta) \geq \alpha$ for all $\theta \in \Theta_1$.

Suppose we want to test $H_0\colon \theta = \theta_0$ against $H_1\colon \theta \neq \theta_0$. Consider a test $\phi(x)$. Let $\beta(\theta) = \int \phi(x) f(x \mid \theta)\,\mathrm{d}x$ denote the power function. If $\beta(\theta)$ is differentiable in θ, then for any unbiased test, we necessarily have $\beta'(\theta_0) = 0$, since $\beta(\theta)$ is minimized at θ_0.

In some two-sided testing problems, there exists a UMP test among all unbiased tests of level α even though there are no UMP tests among all tests of level α. The next example is the most important application:

Example 5. Suppose $X \sim \mathcal{N}(\theta, 1)$. Let $\phi(x)$ denote a test of $\theta = \theta_0$ against $\theta \neq \theta_0$. Let $f(x) = (2\pi)^{-1/2} e^{-\frac{x^2}{2}}$ denote the probability density function (PDF) of standard normal distribution. The power is

$$\beta(\theta) = \int_{\mathbb{R}} \phi(x) f(x - \theta)\,\mathrm{d}x.$$

Differentiating under the integral sign yields

$$\beta'(\theta) = \int_{\mathbb{R}} \phi(x) f(x - \theta)(x - \theta)\,\mathrm{d}x.$$

Fix some $\theta_1 \neq \theta_0$ The most powerful unbiased test against this alternative therefore maximizes

$$\int_{\mathbb{R}} \phi(x) f(x - \theta_1)\,\mathrm{d}x.$$

subject to

$$\int_{\mathbb{R}} \phi(x) f(x - \theta_0)(x - \theta_0)\,\mathrm{d}x = 0, \qquad \int_{\mathbb{R}} \phi(x) f(x - \theta_0)\,\mathrm{d}x = \alpha. \tag{2}$$

We now generalize the Neyman-Pearson lemma to claim that a test of the form

$$\phi(x) = \begin{cases} 1 & \text{if } f(x - \theta_1) > \lambda_1 f(x - \theta_0) + \lambda_2 f(x - \theta_0)(x - \theta_0), \\ 0 & \text{if } f(x - \theta_1) < \lambda_1 f(x - \theta_0) + \lambda_2 f(x - \theta_0)(x - \theta_0). \end{cases}$$

for some λ_1, λ_2 chosen to satisfy the constraints (2) solves this problem (this is like the Neyman-Pearson test, except the "null density" is $\lambda_1 f(x - \theta_0) + \lambda_2 f(x - \theta_0)(x - \theta_0)$). We don't need to consider randomized tests since the distribution of the new "likelihood ratio" is continuous. To prove the claim, consider any other unbiased test ϕ'. Then

$$(\phi - \phi')(f(x-\theta_1) - \lambda_1 f(x-\theta_0) - \lambda_2 f(x-\theta_0)(x-\theta_0)) \geq 0,$$

so integrating wrt x yields

$$\begin{aligned} 0 &\leq E_{\theta_1}[\phi - \phi'] - \lambda_1 E_{\theta_0}[\phi - \phi'] - \lambda_2 \int (\phi(x) - \phi'(x)) f(x-\theta_0)(x-\theta_0)\, \mathrm{d}x \\ &= E_{\theta_1}[\phi - \phi'] - \lambda_1 E_{\theta_0}[\phi - \phi'] \leq E_{\theta_1}[\phi - \phi'], \end{aligned}$$

where the second line follows since the tests are unbiased. This proves the claim that ϕ maximizes power at θ_1 among all unbiased tests. Now, ϕ rejects if $e^{-(x-\theta_1)^2/2} > \lambda_1 e^{-(x-\theta_0)^2/2} + \lambda_2 e^{-(x-\theta_0)^2/2}(x-\theta_0)$, or, equivalently, if

$$e^{xb} > a_0 + a_1 x,$$

where $b = \theta_1 - \theta_0$, $a_0 = e^{\theta_1^2/2 - \theta_0^2/2}(\lambda_1 - \lambda_2\theta_0)$, and $a_1 = e^{\theta_1^2/2 - \theta_0^2/2}\lambda_2$.

There are two cases to consider: $b < 0$ or $b > 0$, depending on whether we fixed θ_1 to be greater or smaller than θ_0. If $b < 0$, and $a_1 \geq 0$, then the rejection region is one-sided (since the left-hand side is decreasing and the right-hand side is non-decreasing), which cannot be the case since such a rejection region doesn't produce an unbiased test. So if $b < 0$, we must have $a_1 < 0$, and the rejection region is a complement of an interval (draw a picture to see this), $\phi(X) = \mathbb{1}\{X \notin [C_1, C_2]\}$. By a symmetric argument, if $b > 0$, we must have $a_1 > 0$, and the rejection region has the same shape. But to satisfy the unbiasedness requirement, it must be the case that the midpoint of the interval $[C_1, C_2]$ is θ_0. To see this, note that $\int (x-\theta_0) f(x-\theta_0)\, \mathrm{d}x = E_{\theta_0}[X - \theta_0] = 0$, so that $\beta'(\theta_0) = \int \mathbb{1}\{x \notin [C_1, C_2]\}(x-\theta_0) f(x-\theta_0)\, \mathrm{d}x = -\int_{C_1}^{C_2}(x-\theta_0) f(x-\theta_0)\, \mathrm{d}x = -\int_{C_1-\theta_0}^{C_2-\theta_0} x f(x)\, \mathrm{d}x$. This equals zero, by a symmetry argument, only if $C_2 - \theta_0 = -(C_1 - \theta_0) \iff (C_1 + C_2)/2 = \theta_0$. So the acceptance region is $[2\theta_0 - C_2, C_2]$. The size constraint gives $C_2 - \theta_0 = z_{1-\alpha/2}$, the $1-\alpha/2$ quantile of a normal distribution. Therefore, the UMP unbiased test rejects when $|X - \theta_0| \geq z_{1-\alpha/2}$. Since this test doesn't depend on the particular alternative θ_1, it must be UMP unbiased against $H_1 \colon \theta \neq \theta_0$.

So, in conclusion, the two-sided z-test is UMP unbiased for testing $H_0 \colon \theta = \theta_0$ against a two-sided alternative. ⊠

REFERENCES

Neyman, Jerzy, and Egon Sharpe Pearson. 1933. "On the Problem of the Most Efficient Tests of Statistical Hypotheses." *Philosophical Transactions of the Royal Society of London*, Series A, 231, nos. 694-706 (February): 289–337. https://doi.org/10.1098/rsta.1933.0009.