STA732
Statistical Inference

Lecture 16: Hypothesis testing

Yuansi Chen

Spring 2022

Duke University

https://www2.stat.duke.edu/courses/Spring22/sta732.01/

Recap from Lecture 15

- Consistency and Asymptotic normality of MLE

Asymptotic distribution of MLE

Thm. 9.14 in Keener

$X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} p_{\theta_0}, \theta_0 \in \Omega$. Assume

- $\hat{\theta}_n \in \arg\max_\theta \ell_n(\theta; X)$, is consistent
- In a neighborhood $\bar{B}_\epsilon(\theta_0) = \left\{\theta : \|\theta - \theta_0\|_2 \le \epsilon\right\}$
    - $\ell_1(\theta; x)$ has two continuous derivatives, $\forall x$
    - $\mathbb{E}\left[\sup_{\theta \in \bar{B}_\epsilon} \left\|\nabla^2 \ell_1(\theta; X_i)\right\|_2\right] < \infty$
- Fisher information $I_1(\theta_0) \succeq 0$

Then

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \Rightarrow \mathcal{N}(0, I_1(\theta_0)^{-1})$$

Recap on point estimation

Typical ways to argue for an optimal estimator in point estimation given model P_θ, data X, estimand $g(\theta)$, loss function L.

- Restrict to samller class: unbiased, equivariant
- Apply global measures of optimality: Bayes, minimax
- Large sample approach: MLE or estimators related to MLE

We have finished the Lehmann and Casella book!

Goal of Lecture 16

1. Hypothesis testing
2. The Neyman-Pearson Paradigm
3. Neyman-Pearson Lemma

Chap. 12.1, 12.2, 12.3, 12.4 of Keener or Chap. 3 of Lehmann and Romano

Hypothesis testing

Suppose $X \sim P_\theta$ for some $\theta \in \Omega$. We divide the model space Ω into two mutually exclusive subsets $\Omega = \Omega_0 \cup \Omega_1$ and $\Omega_0 \cap \Omega_1 = \emptyset$. Consider the following two hypotheses

$$H_0 : \theta \in \Omega_0 \text{ (null hypothesis)}$$
$$H_1 : \theta \in \Omega_1 \text{ (alternative hypothesis)}$$

The decision to make is either

- accept H_0 (fail to reject, no definite conclusion)
- or reject H_0 (conclude H_0 is likely to be false)

Examples

- $X \sim \mathcal{N}(\theta, 1)$

$$H_0 : \theta \leq 0 \text{ vs } H_1 : \theta > 0$$
$$\text{or } H_0 : \theta = 0 \text{ vs } H_1 : \theta \neq 0$$

- $X_1, \dots, X_n \sim P, Y_1, \dots, Y_m \sim Q$

$$H_0 : P = Q \text{ vs } H_1 : P \neq Q$$

Decision rule in hypothesis testing: test function

A decision rule can be expressed as a test function (also critical function) $\phi(X)$ which takes values in $[0, 1]$, defined as

$$\phi(X) = \text{the probability of rejecting } H_0 \text{ given } X.$$

Note the above definition allows the the decision to be randomized

- If $\phi(x)$ takes values in $(0, 1)$ for some x, we way that it is a randomized test
- For a non-randomized test ϕ, the rejection region is

$$\mathcal{R} = \{x : \phi(x) = 1\}$$

The accept region is $\mathcal{A} = \mathbf{S} \backslash \mathcal{R}$

Examples of test function (1)

For x on the boundary of the reject and accept regions, we reject H_0 with probability γ.

$$\phi(x) = \begin{cases} 1 & x \in \mathcal{R} \\ \gamma & x \text{ on the boundary} \\ 0 & x \in \mathcal{A} \end{cases}$$

Examples of test fucntion (2)

$X \sim \mathcal{N}(\theta, 1)$, $H_0 : \theta = 0$, $H_1 : \theta \neq 0$
Let $z_\alpha = \Phi^{-1}(1-\alpha)$, Φ is the normal cdf. Here are some test functions

- $\phi_1(x) = \mathbf{1}_{x > z_\alpha}$, 1-sided test
- $\phi_2(x) = \mathbf{1}_{|x| > z_{\alpha/2}}$, 2-sided test
- $\phi_3(x) = \mathbf{1}_{x < z_{\alpha/3} \text{ or } x > z_{2\alpha/3}}$

Power and level of significance

- The power function of a test $\phi(X)$ is

$$\beta_\phi(\theta) = \mathbb{E}_\theta \phi(X) = P_\theta(X \in \mathcal{R})$$

It is the probabilty that the test rejects the null hypothesis for a given θ

- The significance level of a test $\phi(X)$ is

$$\sup_{\theta \in \Omega_0} \mathbb{E}_\theta(\phi(X)) = \sup_{\theta \in \Omega_0} \beta_\phi(\theta)$$

- We say ϕ is a level-α test if its significance level is $\leq \alpha$

Examples

These are also level-α tests:

$X \sim \mathcal{N}(\theta, 1), H_0 : \theta = 0, H_1 : \theta \neq 0$

Let $z_\alpha = \Phi^{-1}(1-\alpha)$, Φ is the normal cdf. Here are some test functions

- $\phi_1(x) = \mathbf{1}_{x > z_\alpha}$, 1-sided test
- $\phi_2(x) = \mathbf{1}_{|x| > z_{\alpha/2}}$, 2-sided test
- $\phi_3(x) = \mathbf{1}_{x < z_{\alpha/3} \text{ or } x > z_{2\alpha/3}}$

draw the power function

Simple versus simple

A hypothesis is called simple if it completely specifies the distribution of the data. For example, $H_i : \theta \in \Omega_i$ is simple when Ω_i contains a single parameter value θ_i

Simple versus simple testing is the case where both H_0 and H_1 are simple.

A loss function for simple versus simple hypothesis testing

Under simple versus simple testing, consider the loss function

$$L(0, \phi) = \mathbf{1}_{\phi=1}$$
$$L(1, \phi) = c\mathbf{1}_{\phi=0}$$

where $c \geq 0$ constant

Then the risk functions are

$$R(0, \phi) = \mathbb{E}_{\theta_0} L(0, \phi(X)) = \mathbb{P}_{\theta_0}(\phi(X) = 1) = \beta_\phi(\theta_0)$$
$$R(1, \phi) = \mathbb{E}_{\theta_0} cL(1, \phi(X)) = c \cdot \mathbb{P}_{\theta_1}(\phi(X) = 0) = c \cdot (1 - \beta_\phi(\theta_1))$$

When H_0 and H_1 are composite, we can't really write the risk function this way, need to deal with θ

With the above, we are back to the decision theoretic framework:

Ask the question of finding the "optimal" ϕ under the risk function

The Neyman-Pearson Paradigm

The Neyman-Pearson Paradigm

Fix the level of significance α, find the ϕ such that

$$\beta_\phi(\theta) = \mathbb{E}_\theta \phi \text{ is large if } \theta \in \Omega_1.$$

In the simple versus simple testing

We look for a level-α test that has the largest probability of rejecting the null when the truth is alternative (this probabiliy is also called power). In other words

$$\begin{aligned} &\max_\phi \beta_\phi(\theta_1) \\ &\text{s.t. } \beta_\phi(\theta_0) \le \alpha \end{aligned}$$

Example

X	1	2	3	4	5
$P_0(X=i)$	$\frac{1}{4}$	$\frac{1}{100}$	$\frac{1}{100}$	$\frac{3}{100}$	$\frac{7}{10}$
$P_1(X=i)$	$\frac{1}{2}$	$\frac{1}{10}$	$\frac{2}{100}$	$\frac{5}{100}$	$\frac{33}{100}$
$P_1(X=i)/P_0(X=i)$	2	10	2	$5/3$	$33/70$

- Calculate the significance level and power function of the following test

$$\phi(X) = \begin{cases} 1 & \text{if } X \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

- If we set $\alpha = \frac{1}{100}$, what would be an optimal test? (grocery shopper analogy)

Likelihood ratio test

In the simple versus simple testing, a likelihood ratio test is

$$\phi^*(x) = \begin{cases} 1 & \text{if } \frac{p_{\theta_1}(x)}{p_{\theta_0}(x)} > c \\ \gamma & \text{if } \frac{p_{\theta_1}(x)}{p_{\theta_0}(x)} = c \\ 0 & \text{otherwise} \end{cases}$$

The level-α likelihood ratio test (LRT) is the one that chooes c, γ such that

$$\mathbb{E}_{\theta_0} \phi^*(X) = \alpha.$$

Neyman-Pearson Lemma

Neyman-Pearson Lemma

Neyman-Pearson Lemma, Keener Prop 12.2

Given any level $\alpha \in [0, 1]$, there exists a LRT ϕ_α with level α. And any level-α LRT maximizes $\mathbb{E}_{\theta_1} \phi$ among all tests with level at most α.

Neyman-Pearson Lemma

Neyman-Pearson Lemma, Keener Prop 12.2

Given any level $\alpha \in [0, 1]$, there exists a LRT ϕ_α with level α. And any level-α LRT maximizes $\mathbb{E}_{\theta_1} \phi$ among all tests with level at most α.

Prop 12.3 shows that if a test is optimal, then it must be a LRT

The power function has two values in simple vs simple testing

$$\beta_\phi(\theta_0) = \mathbb{E}_{\theta_0}\phi = \int \phi(x) p_{\theta_0}(x) d\mu(x)$$

$$\beta_\phi(\theta_1) = \mathbb{E}_{\theta_1}\phi = \int \phi(x) p_{\theta_1}(x) d\mu(x)$$

Prop 12.1 in Keener

Suppose $k \geq 0$, ϕ^* maximizes

$$\mathbb{E}_{\theta_1}\phi - k\mathbb{E}_{\theta_0}\phi$$

among all test functions, and $\mathbb{E}_{\theta_0}\phi^* = \alpha$. Then ϕ^* maximizes $\mathbb{E}_{\theta_1}\phi$ over all ϕ with level at most α

proof: For ϕ with level at most α, $\mathbb{E}_{\theta_0}\phi(X) \leq \alpha$. Then

$$\begin{aligned}\mathbb{E}_{\theta_1}[\phi(X)] &\leq \mathbb{E}_{\theta_1}[\phi(X)] + k(\alpha - \mathbb{E}_{\theta_0}[\phi(X)]) \\ &\leq \mathbb{E}_{\theta_1}[\phi^*(X)] - k\mathbb{E}_{\theta_0}[\phi^*(X)] + k\alpha \\ &= \mathbb{E}_{\theta_1}[\phi^*(X)]\end{aligned}$$

Proof of the Neyman Pearson Lemma

According to Prop 12.1, it is sufficient to show that likelihood ratio test with level α maximizes the Lagrangian form

$$
\begin{aligned}
&\mathbb{E}_{\theta_1}\phi(X) - k\mathbb{E}_{\theta_0}\phi(X) \\
&= \int (p_{\theta_1}(x) - kp_{\theta_0}(x))\phi(x)d\mu(x) \\
&= \int_{p_{\theta_1} > kp_{\theta_0}} \left|p_{\theta_1} - kp_{\theta_0}\right| \phi d\mu - \int_{p_{\theta_1} < kp_{\theta_0}} \left|p_{\theta_1} - kp_{\theta_0}\right| \phi d\mu
\end{aligned}
$$

To maximize the above expression, ϕ^* must have

$$\phi^*(x) = 1 \text{ when } p_{\theta_1}(x) > kp_{\theta_0}(x)$$

$$\phi^*(x) = 0 \text{ when } p_{\theta_1}(x) < kp_{\theta_0}(x)$$

So the test is based on LR. It remains to set the correct level.

Proof of the Neyman Pearson Lemma (2)

Choose minimum $k \geq 0$, such that

$$\mathbb{P}_{\theta_0}\left[\frac{p_{\theta_1}(X)}{p_{\theta_0}(X)} > k\right] \leq \alpha \leq \mathbb{P}_{\theta_0}\left[\frac{p_{\theta_1}(X)}{p_{\theta_0}(X)} \geq k\right]$$

And choose γ to "top up" the significane level

$$\mathbb{P}_{\theta_0}\left[\frac{p_{\theta_1}(X)}{p_{\theta_0}(X)} > k\right] + \gamma \mathbb{P}_{\theta_0}\left[\frac{p_{\theta_1}(X)}{p_{\theta_0}(X)} = k\right] = \alpha$$

Picture for choosing k_α, γ_α for ϕ^*

Lower bound on the power of LRT

Cor 12.4 in Keener

If $p_{\theta_0} \neq p_{\theta_1}$ and ϕ_α is a level-α likelihood ratio test with $\alpha \in (0, 1)$, then $\mathbb{E}_{\theta_1} \phi_\alpha > \alpha$.

Example: LRT for exponential family

$$X \sim p_\eta(x) = e^{\eta T(x) - A(\eta)} h(x)$$

$H_0 : \eta = \eta_0$ vs $H_1 : \eta = \eta_1 > \eta_0$

doe the optimal test depend on the exact value of η_1?

Summary

- Hypothesis testing is a model choice problem, which can be formulated as a decision theoretic problem
- Neyman-Pearson paradigm took a constrained optimization formulation
- Neyman-Pearson lemma show that likelihood-ratio tests are optimal (in simple vs simple)

What is next?

Uniformly most powerful (UMP) tests

Thank you for attending

See you on Monday in Old Chem 025