STA 522, Spring 2022
Introduction to Theoretical Statistics II

Lecture 10

Department of Biostatistics
University at Buffalo

AGENDA

- Properties of tests, finding c in LRT
- Methods of evaluating tests
- Neyman Pearson Lemma

Review: likelihood ratio test

- Recall the **likelihood function** $L(\theta \mid \underline{x}) = f(\underline{x} \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta)$. The **likelihood ratio test (LRT) statistic** for testing $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta_0^c$ is $\lambda(\underline{x}) = \dfrac{\sup_{\Theta_0} L(\theta \mid \underline{x})}{\sup_{\Theta} L(\theta \mid \underline{x})}$.

- Note that the likelihood ratio test statistic can be viewed as
$$\lambda(\underline{x}) = \frac{L(\hat{\theta}_0 \mid \underline{x})}{L(\hat{\theta} \mid \underline{x})} = \frac{\text{restricted maximization}}{\text{unrestricted maximization}},$$
where $\hat{\theta}$ is the MLE obtained by maximizing $L(\theta \mid \underline{x})$ over the entire parameter space Θ, and $\hat{\theta}_0$ is the MLE obtained by maximizing over the restricted parameter space Θ_0.

- A **likelihood ratio test (LRT)** is any test that has a rejection region of the form $\{\underline{x} : \lambda(\underline{x}) \leq c\}$, where $c \in [0, 1]$.

- Question: how to determine the threshold c? Heuristic idea: want a c such that we won't, or at least very infrequently will, reject H_0 when it is in fact true, or not reject H_0 when it is in fact false.

Errors in Hypothesis Testing

Definition: Suppose we are testing

$$H_0 : \theta \in \Theta_0$$
$$\text{vs. } H_1 : \theta \in \Theta_0^c.$$

If $\theta \in \Theta_0$, but the test incorrectly rejects H_0, then the test has made a **Type I error**.

If, on the other hand, $\theta \in \Theta_0^c$, but the test decides to accept H_0, then the test has made a **Type II error**.

		Decision	
		Accept H_0	Reject H_0
Truth	H_0	Correct decision	Type I Error
	H_1	Type II Error	Correct decision

Computing Error Probabilities

Definition: Let R denote the rejection region of a hypothesis test.

If $\theta \in \Theta_0$, then the probability of a Type I error is

$$P_\theta(\underline{X} \in R).$$

If $\theta \in \Theta_0^c$, then the probability of a Type II error is

$$P_\theta(\underline{X} \notin R) = 1 - P_\theta(\underline{X} \in R).$$

Power Function

Definition: The **power function** of a hypothesis test with rejection region R is the function of θ defined by

$$\beta(\theta) = P_\theta(\underline{X} \in R)$$
$$= \begin{cases} \text{probability of a Type I error} & \text{if } \theta \in \Theta_0 \\ 1 - \text{ probability of a Type II error} & \text{if } \theta \in \Theta_0^c. \end{cases}$$

Comments on the Power function:

(a) Ideally, we want $\beta(\theta) = 0$ for all $\theta \in \Theta_0$ and $\beta(\theta) = 1$ for all $\theta \in \Theta_0^c$.

(b) Depends on the hypothesis test (what are we testing?).

(c) Depends on the rejection region (value of c).

(d) It's a function of θ, not the data.

(e) Since it's a probability, $0 \leq \beta(\theta) \leq 1$ for all θ.

Example: Suppose $X \sim \text{binomial}(5, \theta)$, and we are testing $H_0 : \theta \leq \frac{1}{2}$ vs. $H_1 : \theta > \frac{1}{2}$. Consider the two rejection regions

$$R_1 = \{x : x = 5\}$$

$$R_2 = \{x : x = 3, 4, 5\}.$$

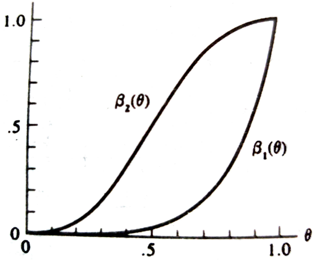
Note that with R_1, we reject H_0 if and only if we observe all successes, whereas with R_2, we reject H_0 if and only if we observe at least 3 successes. Determine the power function for each test.

Here

$$\beta_1(\theta) = P_\theta(X \in R_1) = P_\theta(X = 5) = \binom{5}{5}\theta^5(1-\theta)^{5-5} = \theta^5$$

$$\beta_2(\theta) = P_\theta(X \in R_2) = \sum_{j=3}^{5} P_\theta(X = j) = \sum_{j=3}^{5} \binom{5}{j}\theta^j(1-\theta)^{5-j}$$

Comments about the two power functions



(a) $\beta_2(\theta)$ has higher Type I error and lower Type II error.

(b) $\beta_1(\theta)$ has lower Type I error and higher Type II error.

(c) Ideally, what we will do is try to maximize power while controlling Type I error.

(d) This is how we will choose c in our previous calculations of rejection regions.

Size and Level

Definition: For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a **size** α **test** if

$$\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha.$$

For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a **level** α **test** if

$$\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha.$$

Notes: the set of size α tests is a subset of the set of level α tests.

By specifying the level of a test, we are only controlling the Type I error, not the Type II error.

Choosing c For LRTs

- Restricting to size α tests allows us to determine the value of c to use in the LRT.

- We can build a size α LRT by choosing c so that

$$\sup_{\theta \in \Theta_0} P_\theta(\underline{X} \in R) = \alpha, \quad \text{i.e.,} \quad \sup_{\theta \in \Theta_0} P_\theta(\lambda(\underline{X}) \leq c) = \alpha.$$

Example (contd.): Let $X_1, X_2, \ldots, X_n \sim$ iid $\mathrm{N}(\theta, 1)$. Suppose we wish to test $H_0 : \theta = \theta_0$vs. $H_1 : \theta \neq \theta_0$. We saw that the LRT rejection region is given by

$$R = \{\underline{x} \,:\, |\overline{x} - \theta_0| \geq k\},$$

where $k = \sqrt{\frac{-2\log c}{n}}$. Find the value of c so that we have a size α test.

Since $\Theta_0 = \{\theta_0\}$ is singleton, hence

$$\text{size} = \sup_{\Theta_0} P_\theta \left(|\overline{X} - \theta_0| \geq k\right) = P_{\theta_0} \left(|\overline{X} - \theta_0| \geq k\right)$$

Now, under H_0, $\overline{X} \sim N(\theta_0, 1/n)$ so that $Z = \sqrt{n}(\overline{X} - \theta_0) \sim N(0, 1)$. Therefore the size of the LRT being α implies

$$\begin{aligned}
\alpha &= P_{\theta_0} \left(|\sqrt{n}(\overline{X} - \theta_0)| \geq \sqrt{n}\ k\right) \\
&= P_{\theta_0}(|Z| \geq \sqrt{n}\ k) \\
&= P(Z \geq \sqrt{n}\ k) + P(Z \leq -\sqrt{n}\ k) \\
&= P(Z \geq \sqrt{n}\ k) + P(-Z \geq -\sqrt{n}\ k) = 2\ P(Z \geq \sqrt{n}\ k)
\end{aligned}$$

Let z_α be the upper α-th quantile of Z such that $P(Z \geq z_\alpha) = \alpha$.

Here $\alpha/2 = P(Z \geq \sqrt{n}\ k)$, which implies

$$\sqrt{n}\ k = z_{\alpha/2} \implies k = \frac{1}{\sqrt{n}} z_{\alpha/2} \implies c = \exp\left(-z_{\alpha/2}^2/2\right)$$

Example (contd.): Let $X_1, X_2, \ldots, X_n \sim$ iid from a location exponential population with pdf

$$f(x \mid \theta) = e^{-(x-\theta)} I_{[\theta,\infty)}(x).$$

Suppose we wish to test $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$. We showed that the LRT rejection region is given by

$$R = \left\{ \underline{x} \,:\, x_{(1)} \geq \theta_0 - \frac{\log c}{n} \right\}.$$

Find the value of c so that we have a size α test.

HW. See p. 386 in the textbook.

Evaluating Tests

Definition: A test with power function $\beta(\theta)$ is **unbiased** if

$$\beta(\theta') \geq \beta(\theta'')$$

for every $\theta' \in \Theta_0^c$ and $\theta'' \in \Theta_0$.

Definition: Let $\mathcal{C}$ be a class of tests for testing $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta_0^c$. A test in class $\mathcal{C}$, with power function $\beta(\theta)$, is a **uniformly most powerful (UMP) class $\mathcal{C}$ test** if $\beta(\theta) \geq \beta'(\theta)$ for every $\theta \in \Theta_0^c$ and every $\beta'(\theta)$ that is a power function of a test in class $\mathcal{C}$.

Note: if we take $\mathcal{C}$ to be the class of all level α tests, the test described in the above definition is called a **UMP level α test**.

Neyman-Pearson Lemma

Theorem 8.3.12

Consider testing $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$, where

(1) the pdf or pmf corresponding to θ_i is $f(\underline{x} \,|\, \theta_i)$ for $i = 0, 1$;

(2) the test has a rejection region R that satisfies

$\underline{x} \in R \quad \text{if} \quad f(\underline{x} \,|\, \theta_1) > kf(\underline{x} \,|\, \theta_0)$ and $\underline{x} \in R^c \quad \text{if} \quad f(\underline{x} \,|\, \theta_1) < kf(\underline{x} \,|\, \theta_0)$

for some $k \geq 0$; and

(3) $\alpha = P_{\theta_0}(\underline{X} \in R)$.

Then

(a) **(Sufficiency)** any test that satisfies (2) and (3) above is a UMP level α test; and

(b) **(Necessity)** if there exists a test satisfying (2) and (3) above with $k > 0$, then every UMP level α test is a size α test (satisfies (3) above), and every UMP level α test satisfies (2) above, except perhaps on a set A satisfying $P_{\theta_0}(\underline{X} \in A) = P_{\theta_1}(\underline{X} \in A) = 0$.

Tests Based on Sufficient Statistics

Corollary 8.3.13

Consider testing $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$. Suppose $T(\underline{X})$ is a sufficient statistic for θ, and let $g(t \mid \theta_i)$ be the pdf or pmf of T corresponding to θ_i for $i = 0, 1$. Then any test based on T with rejection region S (a subset of the sample space of T) is a UMP level α test if it satisfies

(1) for some $k \geq 0$,

$$t \in S \quad \text{if} \quad g(t \mid \theta_1) > kg(t \mid \theta_0)$$

and

$$t \in S^c \quad \text{if} \quad g(t \mid \theta_1) < kg(t \mid \theta_0)$$

and

(2) $\alpha = P_{\theta_0}(T \in S)$.

Proof: Use factorization theorem. Reading exercise. See p. 390 in the textbook.

Example: Suppose $X \sim \text{binomial}(2, \theta)$, and we are testing $H_0 : \theta = \frac{1}{2}$ vs. $H_1 : \theta = \frac{3}{4}$. Determine the UMP level α tests for $\alpha = 0, \frac{1}{4}, \frac{3}{4}, 1$.

At the outset note that a "larger" value of X favors H_1, and a smaller value of X favors H_0.

We have $f(x \mid \theta) = \binom{2}{x} \theta^x \, (1-\theta)^{2-x}$; $x = 0, 1, 2$. Consider the ratio

$$\frac{f\left(x \mid \theta = \frac{3}{4}\right)}{f\left(x \mid \theta = \frac{1}{2}\right)} = \frac{\binom{2}{x} \, \left(\frac{3}{4}\right)^x \, \left(\frac{1}{4}\right)^{2-x}}{\binom{2}{x} \, \left(\frac{1}{2}\right)^x \, \left(\frac{1}{2}\right)^{2-x}} = \left(\frac{3}{2}\right)^x \, \left(\frac{1}{2}\right)^{2-x} ; \quad x = 0, 1, 2$$

Therefore,

$$\frac{f\left(0 \mid \theta = \frac{3}{4}\right)}{f\left(0 \mid \theta = \frac{1}{2}\right)} = \frac{1}{4}; \quad \frac{f\left(1 \mid \theta = \frac{3}{4}\right)}{f\left(1 \mid \theta = \frac{1}{2}\right)} = \frac{3}{4}; \quad \frac{f\left(2 \mid \theta = \frac{3}{4}\right)}{f\left(2 \mid \theta = \frac{1}{2}\right)} = \frac{9}{4}.$$

(a) If we choose $\frac{3}{4} < k < \frac{9}{4}$ then NP Lemma says that the test that rejects H_0 if $X = 2$ is the UMP level $\alpha = P\left(X = 2 \mid \theta = \frac{1}{2}\right) = \frac{1}{4}$ test.

(b) If we choose $\frac{1}{4} < k < \frac{3}{4}$ then NP Lemma says that the test that rejects H_0 if $X = 1$ or $X = 2$ is the UMP level $\alpha = P\left(X = 1 \text{ or } 2 \mid \theta = \frac{1}{2}\right) = \frac{3}{4}$ test

(c) Choosing $k < \frac{1}{4}$ or $k > \frac{9}{4}$ produces UMP level 1 or level 0 tests tests respectively.

- If $k = \frac{3}{4}$, NP lemma says that we must reject H_0 when $X = 2$ and accept H_0 but leaves the action for $X = 1$ undetermined.

  – If we accept H_0 for $X = 1$, we get the UMP level $\alpha = \frac{1}{4}$ test as above (case (a)).

  – If we reject H_0 for $X = 1$, we get the UMP level $\alpha = \frac{3}{4}$ test as above (case (b)).

Example (UMP Normal test): Let $X_1, X_2, \dots, X_n \sim$ iid $N(\theta, \sigma^2)$ population, σ^2 known. Consider testing $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$, where $\theta_0 > \theta_1$. Find the UMP test.

The sample mean $\overline{X}$ is a sufficient statistic for θ. So we'll use the corollary of NP lemma with sufficient statistic.

Here

$$g(\overline{x} \mid \theta_1) > k \; g(\overline{x} \mid \theta_0)$$

is equivalent to (HW, use $\theta_1 - \theta_0 < 0$)

$$\overline{x} < \frac{\frac{2\sigma^2 \log k}{n} - (\theta_0^2 - \theta_1^2)}{2(\theta_1 - \theta_0)}$$

i.e., of the form $\overline{x} < c$. Therefore, by the (corollary to) the NP lemma, a test that rejects H_0 when $\overline{x} < c$ is a UMP size α test, where c is obtained from

$$\alpha = P_{\theta_0}(\overline{X} < c) = P_{\theta_0}\left(\frac{\overline{X} - \theta_0}{\sigma/\sqrt{n}} < \frac{c - \theta_0}{\sigma/\sqrt{n}}\right) \implies \frac{c - \theta_0}{\sigma/\sqrt{n}} = z_{1-\alpha} = -z_\alpha$$

i.e., $c = \theta_0 - z_\alpha \frac{\sigma}{\sqrt{n}}$

Comments:

- NP lemma can handle only tests with a point null against a point alternative
- Question: can we say something similar more general hypothesis tests?
- Answer: Yes, to some extent. Will need monotone likelihood ratio & Karlin-Rubin theorem. Next week..