STAT 713 Feb 28, 2019

Chapter 8 Hypo. Testing

Section 8.1 Introduction

- A hypothesis is a statement about a population parameter.

- In hypothesis testing, we have two hypothesis's denoted by H_0 and H_1 (H_a)
  - H_0 ↓ the null hypothesis
  - H_1 ↓ the alternative hypothesis

Test H_0 versus H_1

Suppose the parameter of interest is θ, of which the parameter space is denoted by Θ

eg:

simple hypothesis: $H_0: \theta = \theta_0$ vs. $H_1: \theta \neq \theta_0$ where θ_0 is given

composite hypothesis: $H_0: \theta \leq \theta_0$ vs. $H_1: \theta > \theta_0$

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

General: $H_0: \theta \in \Theta_0$ vs. $H_1: \theta \in \Theta_1$ where $\Theta_0 \cap \Theta_1 = \phi$

test function
$$\phi(\underset{\sim}{x}) = \begin{cases} 1 & \underset{\sim}{x} \in R \\ 0 & \underset{\sim}{x} \notin R \end{cases}$$

8 Hypothesis Testing

Complementary reading: Chapter 8 (CB).

8.1 Introduction

Setting: We observe $\mathbf{X} = (X_1, X_2, ..., X_n) \sim f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k$. For example, $X_1, X_2, ..., X_n$ might constitute a random sample (iid sample) from a population $f_X(x|\boldsymbol{\theta})$. We regard $\boldsymbol{\theta}$ as fixed and unknown.

Definition: A **statistical hypothesis** is a statement about $\boldsymbol{\theta}$. This statement specifies a collection of distributions that $\mathbf{X}$ can possibly have. Two complementary hypotheses in a testing problem are the **null hypothesis**

$$H_0 : \boldsymbol{\theta} \in \Theta_0$$

and the **alternative hypothesis**

$$H_1 : \boldsymbol{\theta} \in \Theta_0^c,$$

where $\Theta_0^c = \Theta \setminus \Theta_0$. We call Θ_0 the **null parameter space** and Θ_0^c the **alternative parameter space**.

Example 8.1. Suppose $X_1, X_2, ..., X_n$ are iid $\mathcal{N}(\theta, \sigma_0^2)$, where $-\infty < \theta < \infty$ and σ_0^2 is known. Consider testing

$$H_0 : \theta = \theta_0$$
versus
$$H_1 : \theta \neq \theta_0,$$

where θ_0 is a specified value of θ. The null parameter space $\Theta_0 = \{\theta_0\}$, a singleton. The alternative parameter space $\Theta_0^c = \mathbb{R} \setminus \{\theta_0\}$.

Terminology: In Example 8.1, we call $H_0 : \theta = \theta_0$ a **simple** (or **sharp**) hypothesis. Note that H_0 specifies exactly one distribution, namely, $\mathcal{N}(\theta_0, \sigma_0^2)$. A simple hypothesis specifies a single distribution.

Terminology: In Example 8.1, suppose we wanted to test

$$H_0 : \theta \leq \theta_0$$
versus
$$H_1 : \theta > \theta_0.$$

We call H_0 a **composite** (or **compound**) hypothesis. Note that H_0 specifies a family of distributions, namely, $\{\mathcal{N}(\theta, \sigma_0^2) : \theta \leq \theta_0\}$.

Goal: In a statistical hypothesis testing problem, we decide between the two complementary hypotheses H_0 and H_1 on the basis of observing $\mathbf{X} = \mathbf{x}$. In essence, a hypothesis test is a specification of the **test function**

$$\phi(\mathbf{x}) = P(\text{Reject } H_0 | \mathbf{X} = \mathbf{x}).$$

Terminology: Let $\mathcal{X}$ denote the support of $\mathbf{X}$.

- The subset of $\mathcal{X}$ for which H_0 is rejected is called the **rejection region**, denoted by R.
- The subset of $\mathcal{X}$ for which H_0 is not rejected is called the **acceptance region**, denoted by R^c.

If

$$\phi(\mathbf{x}) = I(\mathbf{x} \in R) = \begin{cases} 1, & \mathbf{x} \in R \\ 0, & \mathbf{x} \in R^c, \end{cases}$$

the test is said to be **non-randomized**.

Example 8.2. Suppose $X \sim b(10, \theta)$, where $0 < \theta < 1$, and consider testing

$$\begin{gathered} H_0 : \theta \geq 0.35 \\ \text{versus} \\ H_1 : \theta < 0.35. \end{gathered}$$

Here is an example of a **randomized test function**:

$$\phi(x) = \begin{cases} 1, & x \leq 2 \\ \frac{1}{5}, & x = 3 \\ 0, & x \geq 4. \end{cases}$$

Using this test function, we would reject H_0 if $x = 0, 1,$ or 2. If $x = 3$, we would reject H_0 with probability $1/5$. If $x \geq 4$, we would not reject H_0.

- If we observed $x = 3$, we could then subsequently generate $U \sim \mathcal{U}(0, 1)$.
  - If $u \leq 0.2$, then reject H_0.
  - If $u > 0.2$, then do not reject H_0.

Remark: In most problems, a test function ϕ depends on $\mathbf{X}$ through a one-dimensional **test statistic**, say

$$W = W(\mathbf{X}) = W(X_1, X_2, ..., X_n).$$

1. We would like to work with test statistics that are sensible and confer tests with nice statistical properties (does sufficiency play a role?)

2. We would like to find the **sampling distribution** of W under H_0 and H_1.

Example 8.3. Suppose $X_1, X_2, ..., X_n$ are iid $\mathcal{N}(\mu, \sigma^2)$, where $-\infty < \mu < \infty$ and $\sigma^2 > 0$; i.e., both parameters are unknown. Consider testing

$$H_0 : \sigma^2 = 40 \quad \text{versus} \quad H_1 : \sigma^2 \neq 40.$$

In this problem, both

$$\begin{aligned} W_1 = W_1(\mathbf{X}) &= |S^2 - 40| \\ W_2 = W_2(\mathbf{X}) &= \frac{(n-1)S^2}{40} \end{aligned}$$

are reasonable test statistics.

- Because S^2 is an unbiased estimator of σ^2, large values of W_1 (intuitively) are evidence against H_0. However, what is W_1's sampling distribution?
- The advantage of working with W_2 is that we know its sampling distribution when H_0 is true; i.e., $W_2 \sim \chi^2_{n-1}$. It is also easy to calculate the sampling distribution of W_2 when H_0 is not true; i.e., for values of $\sigma^2 \neq 40$.

Example 8.4. McCann and Tebbs (2009) summarize a study examining perceived unmet need for dental health care for people with HIV infection. Baseline in-person interviews were conducted with 2,864 HIV infected individuals (aged 18 years and older) as part of the HIV Cost and Services Utilization Study. Define

$$\begin{aligned} X_1 &= \text{number of patients with private insurance} \\ X_2 &= \text{number of patients with medicare and private insurance} \\ X_3 &= \text{number of patients without insurance} \\ X_4 &= \text{number of patients with medicare but no private insurance.} \end{aligned}$$

Set $\mathbf{X} = (X_1, X_2, X_3, X_4)$ and model $\mathbf{X} \sim \text{mult}(2864, p_1, p_2, p_3, p_4; \sum_{i=1}^4 p_i = 1)$. Under this assumption, consider testing

$$H_0 : p_1 = p_2 = p_3 = p_4 = \tfrac{1}{4} \quad \text{versus} \quad H_1 : H_0 \text{ not true.}$$

Note that an observation like $\mathbf{x} = (0, 0, 0, 2864)$ should lead to a rejection of H_0. An observation like $\mathbf{x} = (716, 716, 716, 716)$ should not. What about $\mathbf{x} = (658, 839, 811, 556)$? Can we find a reasonable one-dimensional test statistic?

8.2 Methods of Finding Tests

Preview: The authors present three methods of finding tests:

1. Likelihood ratio tests (LRTs)
2. Bayesian tests
3. Union-Intersection and Intersection-Union tests (UIT/IUT)

We will focus largely on LRTs. We will discuss Bayesian tests briefly.

8.2.1 Likelihood ratio tests

Recall: Suppose $\mathbf{X} = (X_1, X_2, ..., X_n) \sim f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k$. The **likelihood function** is

$$\begin{aligned} L(\boldsymbol{\theta}|\mathbf{x}) &= f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta}) \\ &\stackrel{\text{iid}}{=} \prod_{i=1}^{n} f_X(x_i|\boldsymbol{\theta}), \end{aligned}$$

where $f_X(x|\boldsymbol{\theta})$ is the common population distribution (in the iid case). Recall that Θ is the **parameter space**.

Definition: The **likelihood ratio test** (**LRT**) **statistic** for testing

$$\begin{gathered} H_0 : \boldsymbol{\theta} \in \Theta_0 \\ \text{versus} \\ H_1 : \boldsymbol{\theta} \in \Theta \setminus \Theta_0 \end{gathered}$$

is defined by

$$\lambda(\mathbf{x}) = \frac{\sup_{\boldsymbol{\theta} \in \Theta_0} L(\boldsymbol{\theta}|\mathbf{x})}{\sup_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}|\mathbf{x})}.$$

A LRT is a test that has a rejection region of the form

$$R = \{\mathbf{x} \in \mathcal{X} : \lambda(\mathbf{x}) \leq c\},$$

where $0 \leq c \leq 1$.

Intuition: The numerator of $\lambda(\mathbf{x})$ is the largest the likelihood function can be over the null parameter space Θ_0. The denominator is the largest the likelihood function can be over the entire parameter space Θ. Clearly,

$$0 \leq \lambda(\mathbf{x}) \leq 1.$$