Lecture 11: Introduction to Hypothesis Testing

Mathematical Statistics II, MATH 60062/70062

Tuesday March 1, 2022

Reference: Casella & Berger, 8.1-8.2.1

Hypothesis testing problem

We observe $\boldsymbol{X} = (X_1, \ldots, X_n) \sim f_{\boldsymbol{X}}(\boldsymbol{x} \mid \boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta$.

A **statistical hypothesis** is a statement about a population parameter $\boldsymbol{\theta}$. This statement specifies a collection of possible values of $\boldsymbol{\theta}$, i.e., the collection of distributions that $\boldsymbol{X}$ can possibly have.

Hypothesis testing problem

We observe $\boldsymbol{X} = (X_1, \ldots, X_n) \sim f_{\boldsymbol{X}}(\boldsymbol{x} \mid \boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta$.

A **statistical hypothesis** is a statement about a population parameter $\boldsymbol{\theta}$. This statement specifies a collection of possible values of $\boldsymbol{\theta}$, i.e., the collection of distributions that $\boldsymbol{X}$ can possibly have.

In a hypothesis testing problem, two complementary hypotheses are called the **null hypothesis** (H_0) and the **alternative hypothesis** (H_1). Typically, we write

$$H_0 : \boldsymbol{\theta} \in \Theta_0 \quad \text{versus} \quad H_1 : \boldsymbol{\theta} \in \Theta_0^c$$

where Θ_0 is some subset of the parameter space and $\Theta_0^c = \Theta \setminus \Theta_0$ is its complement.

Normal hypothesis testing

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

Two common hypothesis testing examples are

- $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, where θ_0 is a specified value of θ. In this case, H_0 specifies a single distribution $\mathcal{N}(\theta_0, \sigma_0^2)$, and is called a **simple** (or **sharp**) hypothesis.
- $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$. In this case, H_0 specifies a family of distributions $\{\mathcal{N}(\theta, \sigma_0^2) : \theta \leq \theta_0\}$, and is called a **composite** (or **compound**) hypothesis.

Hypothesis testing procedure

A **hypothesis testing procedure** or **hypothesis test** is a rule that specifies

1. For which sample values the decision is made to accept H_0 as true.
2. For which sample values H_0 is rejected and H_1 is accepted as true.

The subset of the sample space for which H_0 will be rejected is called the **rejection region** or **critical region**. The complement of the rejection region is called the **acceptance region**.

Hypothesis testing procedure

A **hypothesis testing procedure** or **hypothesis test** is a rule that specifies

1. For which sample values the decision is made to accept H_0 as true.
2. For which sample values H_0 is rejected and H_1 is accepted as true.

The subset of the sample space for which H_0 will be rejected is called the **rejection region** or **critical region**. The complement of the rejection region is called the **acceptance region**.

Typically, a hypothesis test is specified in terms of a **test statistic** $W(X_1, \ldots, X_n) = W(\boldsymbol{X})$, a function of the sample.

Normal hypothesis testing

Suppose $X_1, \ldots, X_n$ are iid $\mathcal{N}(\theta, \sigma^2)$, where $-\infty < \theta < \infty$ and $\sigma^2 > 0$. Both parameters are unknown.

Consider testing

$$H_0 : \sigma^2 = 10 \quad \text{versus} \quad H_1 : \sigma^2 \neq 10.$$

Normal hypothesis testing

Suppose $X_1, \ldots, X_n$ are iid $\mathcal{N}(\theta, \sigma^2)$, where $-\infty < \theta < \infty$ and $\sigma^2 > 0$. Both parameters are unknown.

Consider testing

$$H_0 : \sigma^2 = 10 \quad \text{versus} \quad H_1 : \sigma^2 \neq 10.$$

In this problem,

$$\begin{aligned} W_1 &= |S^2 - 10| \\ W_2 &= \frac{(n-1)S^2}{\sigma^2} \end{aligned}$$

are both reasonable test statistics.

Normal hypothesis testing

Suppose $X_1, \ldots, X_n$ are iid $\mathcal{N}(\theta, \sigma^2)$, where $-\infty < \theta < \infty$ and $\sigma^2 > 0$. Both parameters are unknown.

Consider testing

$$H_0 : \sigma^2 = 10 \quad \text{versus} \quad H_1 : \sigma^2 \neq 10.$$

In this problem,

$$W_1 = |S^2 - 10|$$
$$W_2 = \frac{(n-1)S^2}{\sigma^2}$$

are both reasonable test statistics.

An advantage of working with W_2 is that we know its sampling distribution when H_0 is true.

Hypothesis testing methods

Methods of finding tests

- Likelihood ratio tests
- Bayesian tests

Methods of evaluating tests

- Error probability
- Power function
- Size and level
- Most powerful tests

Likelihood ratio tests

Recall: Suppose that $\boldsymbol{X} = (X_1, \ldots, X_n)$ is a random sample from $f_{\boldsymbol{X}}(\boldsymbol{x} \mid \boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta$. The **likelihood function** is

$$L(\boldsymbol{\theta} \mid \boldsymbol{x}) = f_{\boldsymbol{X}}(\boldsymbol{x} \mid \boldsymbol{\theta}) = \prod_{i=1}^{n} f_X(x_i \mid \boldsymbol{\theta}),$$

where $f_X(x \mid \boldsymbol{\theta})$ is the common population distribution.

Likelihood ratio tests

Recall: Suppose that $\boldsymbol{X} = (X_1, \ldots, X_n)$ is a random sample from $f_{\boldsymbol{X}}(\boldsymbol{x} \mid \boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta$. The **likelihood function** is

$$L(\boldsymbol{\theta} \mid \boldsymbol{x}) = f_{\boldsymbol{X}}(\boldsymbol{x} \mid \boldsymbol{\theta}) = \prod_{i=1}^{n} f_X(x_i \mid \boldsymbol{\theta}),$$

where $f_X(x \mid \boldsymbol{\theta})$ is the common population distribution.

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

$$H_0 : \boldsymbol{\theta} \in \Theta_0 \quad \text{versus} \quad H_1 : \boldsymbol{\theta} \in \Theta_0^c$$

is

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

Likelihood ratio tests

Recall: Suppose that $\boldsymbol{X} = (X_1, \dots, X_n)$ is a random sample from $f_{\boldsymbol{X}}(\boldsymbol{x} \mid \boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta$. The **likelihood function** is

$$L(\boldsymbol{\theta} \mid \boldsymbol{x}) = f_{\boldsymbol{X}}(\boldsymbol{x} \mid \boldsymbol{\theta}) = \prod_{i=1}^{n} f_X(x_i \mid \boldsymbol{\theta}),$$

where $f_X(x \mid \boldsymbol{\theta})$ is the common population distribution.

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

$$H_0 : \boldsymbol{\theta} \in \Theta_0 \quad \text{versus} \quad H_1 : \boldsymbol{\theta} \in \Theta_0^c$$

is

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

An LRT is any test that has a rejection region of the form $R = \{\boldsymbol{x} : \lambda(\boldsymbol{x}) \leq c\}$, where c is any number satisfying $0 \leq c \leq 1$.

- The numerator of $\lambda(\boldsymbol{x})$ is

$$\sup_{\boldsymbol{\theta} \in \Theta_0} L(\boldsymbol{\theta} \mid \boldsymbol{x}) = L(\hat{\boldsymbol{\theta}}_0 \mid \boldsymbol{x}),$$

  where $\hat{\boldsymbol{\theta}}_0$ is the **restricted MLE** over Θ_0.

- The denominator of $\lambda(\boldsymbol{x})$ is

$$\sup_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta} \mid \boldsymbol{x}) = L(\hat{\boldsymbol{\theta}} \mid \boldsymbol{x}),$$

  where $\hat{\boldsymbol{\theta}}$ is the **unrestricted MLE**.

Obviously, $0 \leq \lambda(\boldsymbol{x}) \leq 1$. The ratio is *small* if there are parameter points in Θ_0^c for which the observed sample $\boldsymbol{x}$ is much more likely than for any point in Θ_0. In this situation, the LRT criterion says H_0 *should be rejected*.