Statistics with R

Chapter 5: Hypothesis testing

Tabea Rebafka

October 2018

Master AIMS 2018–19

Outline

Definitions I

Problem

- Let $\{\mathbb{P}_\theta, \theta \in \Theta\}$ be a statistical model and consider two disjoint subsets Θ_0 and Θ_1 of Θ.
- We observe a realization $\mathbf{x}$ of the distribution $\mathbb{P}_\theta$ with unknown θ.
- We want to decide whether $\theta \in \Theta_0$ or $\theta \in \Theta_1$.

Example: Tossing a coin

- Based on the outcomes of n tosses, we would like to decide whether the coin is a fair coin or not.
- Mathematically speaking: Is $p = \frac{1}{2}$ or $p \neq \frac{1}{2}$?

Definitions II

Hypotheses

- $H_0 : \theta \in \Theta_0$ is called the **null hypothesis**.
- $H_1 : \theta \in \Theta_1$ is called the **alternative hypothesis**.
- The null hypothesis (or the alternative) is **simple** if $\Theta_0 = \{\theta_0\}$ (or $\Theta_1 = \{\theta_1\}$). Otherwise the hypothesis is **multiple**.

Definitions III

Test

- A **test** of hypothesis H_0 versus H_1, is a (measurable) function φ that associates to each possible dataset $\mathbf{x}$ a decision.
- More precisely, $\varphi : \mathcal{X} \mapsto \{0, 1\}$ (where $\mathcal{X}$ is the observation space) such that
  - $\varphi(\mathbf{X}) = 0$ means that we **conserve the null hypothesis** H_0 or **do not reject** H_0
  - $\varphi(\mathbf{X}) = 1$ means that we **reject** H_0 or **decide/accept** H_1.
- The **critical region** of test φ is defined by

$$\mathcal{R}_\varphi = \{\mathbf{x} \in \mathcal{X} : \varphi(\mathbf{x}) = 1\}.$$

Each test is completed characterized by it critical region.

Definitions IV

- In general, the critical region $\mathcal{R}_\varphi$ can be naturally written by using some **test statistic $T(\mathbf{X})$** and a set R such that

$$\mathcal{R}_\varphi = \{\mathsf{x} \in \mathcal{X} : T(\mathbf{X}) \in R\}.$$

- Most often, we have

$$\mathcal{R}_\varphi = \{\mathsf{x} \in \mathcal{X} : T(\mathbf{X}) > c\} \quad \text{or} \quad \mathcal{R}_\varphi = \{\mathsf{x} \in \mathcal{X} : T(\mathbf{X}) < c\},$$

for some constant $c \in \mathbb{R}$.

Definitions V

When we decide between H_0 and H_1, there are two ways to make a mistake:

Error types

- **Type 1 error**: we reject H_0 while it is true.
  The probability to commit a type 1 error is the **type 1 error rate** defined as
  $$\alpha(\theta_0) = \mathbb{P}_{\theta_0}(T(\mathbf{X}) \in R) \quad \forall \theta_0 \in \Theta_0.$$
- **Type 2 error**: We conserve H_0 while it is false.
  The probability to commit a type 2 error is the **type 2 error rate** defined as
  $$\beta(\theta_1) = \mathbb{P}_{\theta_1}(T(\mathbf{X}) \notin R) \quad \forall \theta_1 \in \Theta_1.$$

Definitions VI

Significance level

- The **size of the test** α^* is defined as

$$\alpha^* = \sup_{\theta_0 \in \Theta_0} \alpha(\theta_0).$$

- Let $\alpha \in]0,1[$ be fixed. The test has **(significance) level** α if $\alpha^* \leq \alpha$.

Definitions VII

Definition p-value

For every $\alpha \in]0,1[$, let $\mathcal{R}_\alpha = \{\mathbf{x} \in \mathcal{X} : T(\mathbf{x}) \in R_\alpha\}$ be a test with level α of H_0 versus H_1.
Then for a given observation $\mathbf{x}$, the ***p*-value** is the smallest significance level at which H_0 is rejected. That is, the p-value associated to $\mathbf{x}$ is defined by

$$p(\mathbf{x}) = \inf\{\alpha \in]0,1[: T(\mathbf{x}) \in R_\alpha\}.$$

Interpretation of p-values:

- $p(\mathbf{x}) < 0.01$: H_0 is rejected at all usual significance levels. That is the test is in favour of H_1 and we say that the test is **significant**.
- $p(\mathbf{x}) > 0.1$: H_0 cannot be rejected. Either H_0 is true or we do not have enough information to see from the data that H_0 is false.

Definitions VIII

Power

- The **power of a test** is the probability to reject H_1 correctly. The power is defined as

$$\pi(\theta_1) = \mathbb{P}_{\theta_1}(T(\mathbf{X}) \in R), \quad \theta \in \Theta_1.$$

- Let φ_1 and φ_2 be two tests of size α and power π_1 and π_2, respectively. Test φ_1 is **more powerful** than test φ_2 if

$$\pi_1(\theta) \geq \pi_2(\theta) \quad \forall \theta \in \Theta_1.$$

- Let φ^* be a test of size α. We say that the test φ^* is **uniformly most powerful** (UPP) if φ^* is more powerful than any other test of size α.

Construction of tests I

Using estimators

If an estimator $\hat{\theta}$ of θ is known, a natural test consists in rejecting H_0 when $\hat{\theta}$ takes values "close" to Θ_1.

Construction of tests II

Exercise

- Let $X_1, \ldots, X_n$ be i.i.d. observations from the normal distribution $\mathcal{N}(\mu, \sigma^2)$ with unknown $\mu \in \mathbb{R}$ and known $\sigma^2 > 0$.
- Let μ_0 be given.

1. Consider the hypotheses

   $$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu \neq \mu_0.$$

   Construct a test with significance level α. (The test is **two-sided**).

2. Consider the hypotheses

   $$H_0 : \mu \leq \mu_0 \quad \text{versus} \quad H_1 : \mu > \mu_0.$$

   Construct a test with significance level α. (The test is **one-sided**).

Construction of tests III

Using confidence intervals

- Let IC_α be a confidence interval at level $1-\alpha$ for θ.
- Then a test of size α is given by

$$\varphi(\mathbf{x}) = \mathbb{1}\{\Theta_0 \cap IC_\alpha(\mathbf{x}) = \emptyset\} = \mathbb{1}\{\theta_0 \notin IC_\alpha(\mathbf{x}), \forall \theta_0 \in \Theta_0\}.$$

Tests for the normal distribution I

Definition

- Let $X_1, \ldots, X_p$ be i.i.d. random variables with standard normal distribution $\mathcal{N}(0,1)$. Then the random variable $Y = \sum_{i=1}^{p} X_i^2$ has **chi-squared distribution** with p degrees of freedom, and we denote $Y \sim \chi_p^2$.
- Let $U \sim \mathcal{N}(0,1)$, $V \sim \chi_q^2$ two independent random variables. The **Student's *t* distribution** with q degrees of freedom is the distribution of the random variable

$$Y = \frac{U}{\sqrt{V/q}}.$$

  We write $Y \sim t_q$.

Tests for the normal distribution II

Theorem

Let $X_1, \ldots, X_n$ be i.i.d. random variables with normal distribution $\mathcal{N}(\mu, \sigma^2)$ with $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ unknown. Then

1. The sample mean $\bar{X}_n$ has distribution $\mathcal{N}(\mu, \sigma^2/n)$.
2. The sample variance $s_X^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ verifies
$$\frac{n s_X^2}{\sigma^2} \sim \chi^2_{n-1}.$$
3. The sample mean and the sample variance are independent.
4. The statistic
$$\sqrt{n-1}\,\frac{\bar{X}_n - \mu}{s_X}$$
has Student's t-distribution with $n-1$ degrees of freedom.

Tests for the normal distribution III

Exercise

- Let $X_1, \ldots, X_n$ be i.i.d. random variables with normal distribution $\mathcal{N}(\mu, \sigma^2)$ with $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ unknown.
- Let μ_0 be given.

1. Consider the hypotheses

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu \neq \mu_0.$$

   Construct a test with significance level α. (The test is **two-sided**).

2. Consider the hypotheses

$$H_0 : \mu \leq \mu_0 \quad \text{versus} \quad H_1 : \mu > \mu_0.$$

   Construct a test with significance level α. (The test is **one-sided**).