Chapter 5: Hypothesis testing

Tabea Rebafka

October 2018

Master AIMS 2018–19
Outline

1. Definitions

2. Construction of tests

3. Tests for the normal distribution
Problem

- Let $\{P_\theta, \theta \in \Theta\}$ be a statistical model and consider two disjoint subsets $\Theta_0$ and $\Theta_1$ of $\Theta$.
- We observe a realization $x$ of the distribution $P_\theta$ with unknown $\theta$.
- 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 $p \neq \frac{1}{2}$?
## 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**.
A test of hypothesis \( H_0 \) versus \( H_1 \), is a (measurable) function \( \varphi \) that associates to each possible dataset \( x \) a decision.

More precisely, \( \varphi : \mathcal{X} \mapsto \{0, 1\} \) (where \( \mathcal{X} \) is the observation space) such that

- \( \varphi(X) = 0 \) means that we conserve the null hypothesis \( H_0 \) or do not reject \( H_0 \)
- \( \varphi(X) = 1 \) means that we reject \( H_0 \) or decide/accept \( H_1 \).

The critical region of test \( \varphi \) is defined by

\[
R_\varphi = \{ x \in \mathcal{X} : \varphi(x) = 1 \}.
\]

Each test is completed characterized by it critical region.
In general, the critical region $R_\varphi$ can be naturally written by using some test statistic $T(X)$ and a set $R$ such that

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

Most often, we have

$$R_\varphi = \{ x \in \mathcal{X} : T(X) > c \} \text{ or } R_\varphi = \{ x \in \mathcal{X} : T(X) < c \},$$

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:

<table>
<thead>
<tr>
<th>Error types</th>
<th>Description</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type 1 error</strong></td>
<td>We reject $H_0$ while it is true.</td>
<td>$\alpha(\theta_0) = \mathbb{P}_{\theta_0}(T(X) \in R) \quad \forall \theta_0 \in \Theta_0$.</td>
</tr>
<tr>
<td><strong>Type 2 error</strong></td>
<td>We conserve $H_0$ while it is false.</td>
<td>$\beta(\theta_1) = \mathbb{P}_{\theta_1}(T(X) \notin R) \quad \forall \theta_1 \in \Theta_1$.</td>
</tr>
</tbody>
</table>
Significance level

- The **size of the test** $\alpha^*$ 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** $\alpha$ if $\alpha^* \leq \alpha$. 
Definitions VII

Definition $p$-value

For every $\alpha \in ]0, 1[$, let $R_\alpha = \{x \in \mathcal{X} : T(x) \in R_\alpha\}$ be a test with level $\alpha$ of $H_0$ versus $H_1$.

Then for a given observation $x$, the $p$-value is the smallest significance level at which $H_0$ is rejected. That is, the $p$-value associated to $x$ is defined by

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

Interpretation of $p$-values:

- $p(x) < 0.01$: $H_0$ 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**.
- $p(x) > 0.1$: $H_0$ 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.
**Power**

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

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

- Let $\varphi_1$ and $\varphi_2$ be two tests of size $\alpha$ and power $\pi_1$ and $\pi_2$, respectively. Test $\varphi_1$ is **more powerful** than test $\varphi_2$ if

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

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

Using estimators

If an estimator $\hat{\theta}$ of $\theta$ is known, a natural test consists in rejecting $H_0$ when $\hat{\theta}$ takes values “close” to $\Theta_1$. 
Construction of tests II

Exercise

1. 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 \( \sigma^2 > 0 \).

2. Let \( \mu_0 \) be given.

Consider the hypotheses

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

Construct a test with significance level \( \alpha \). (The test is \textit{two-sided}).

Consider the hypotheses

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

Construct a test with significance level \( \alpha \). (The test is \textit{one-sided}).
Construction of tests III

Using confidence intervals

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

$$\varphi(x) = 1\{\Theta_0 \cap IC_\alpha(x) = \emptyset\} = 1\{\theta_0 \notin IC_\alpha(x), \forall \theta_0 \in \Theta_0\}.$$
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 $Y \sim \chi^2_p$.

- Let $U \sim \mathcal{N}(0, 1)$, $V \sim \chi^2_q$ 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 $s^2_X = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$ verifies

$$\frac{ns^2_X}{\sigma^2} \sim \chi^2_{n-1}.$$ 

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 $\mu_0$ 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 $\alpha$. (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 $\alpha$. (The test is one-sided).