Imperial College London

Lecture 09: Hypothesis Testing

Statistical Modelling I

Dr. Riccardo Passeggeri

Outline

Introduction

Motivation

- **(Point) estimator**: one number only (does not reflect uncertainty)
- **Confidence interval**: random interval that contains the true parameter with a certain probability
- **Hypothesis test**: decision rule to choose between one of two statements about the true parameter

Definition: Null Hypothesis and Alternative Hypothesis, Hypothesis Test and Rejection Region

- The two complementary hypotheses in a hypothesis testing problem are called the **null hypothesis** and the **alternative hypothesis**, denoted by H_0 and H_1, respectively.
- A **hypothesis test** is a rule that specifies for which values of the sample $X_1, \ldots, X_n$ the decision is made to accept H_0 as true and for which values to reject H_0 and accept H_1 as true.
- The **rejection region** or **critical region** is a set of values for the test statistic for which H_0 is rejected. *i.e.* if the observed test statistic is in the critical region then we reject H_0 and accept H_1.

Two Types of Errors

	H_0 true	H_0 false
do not reject H_0	✓	Type II error
reject H_0	Type I error	✓

Level of a test: A test is of level α $(0 < \alpha < 1)$ if

$$P_\theta(\text{reject } H_0) \leq \alpha \quad \forall \theta \in \Theta_0.$$

Usually α is small, e.g. 0.01 or 0.05.
Loosely speaking: the probability of a type I error is less than α.
There is no such bound for the probability of a type II error.

Power of a Test

Definition: Power

Setup: Θ parameter space, $\Theta_0 \subset \Theta$, $\Theta_1 = \Theta \setminus \Theta_0$. Consider

$$H_0 : \theta \in \Theta_0 \text{ v.s. } H_1 : \theta \in \Theta_1$$

Suppose we have some test for this hypothesis.
The *power function* is defined as the mapping

$$\beta : \Theta \to [0, 1], \beta(\theta) = P_\theta(\text{reject } H_0)$$

If $\theta \in \Theta_0$ then we want $\beta(\theta)$ to be small.
If $\theta \in \Theta_1$ then we want $\beta(\theta)$ to be large.

Example: $X \sim N(\theta, 1)$, $\theta \in \mathbb{R}$ unknown

$$H_0 : \theta \leq 0 \quad \text{against} \quad H_1 : \theta > 0$$

Level α test

$$\Theta = \mathbb{R}, \Theta_0 = (-\infty, 0], \Theta_1 = (0, \infty)$$

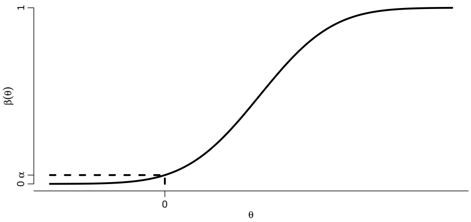
Rejection region

$$R = [c, \infty)$$

Choose c s.t. $\Phi(c) = 1 - \alpha$. Then

$$P_\theta(\text{reject } H_0) = P_\theta(X \geq c)$$

Power of the test



The p-value

Definition

Often the so-called p-value is reported (instead of a test decision):

$$p = \sup_{\theta \in \Theta_0} P_\theta(\text{observing something ''at least as extreme'' as the observed values})$$

Reject H_0 iff $p \leq \alpha \rightarrow \alpha$-level test.
If the test is based on the statistic T with rejection for large values of T then

$$p = \sup_{\theta \in \Theta_0} P_\theta(T \geq t),$$

where t is the observed value.
In the previous example (where $X \sim N(\theta, 1)$ and $H_0 : \theta \leq 0$ against $H_1 : \theta > 0$) the p-value is:

$$p = \sup_{\theta \in \Theta_0} P_\theta(X \geq x) = P_0(X \geq x) = 1 - \Phi(x)$$

Example: $X_1, \ldots, X_n \sim N(\mu, 1)$ iid, μ unknown

$$H_0 : \mu = \mu_0 \text{ against } H_1 : \mu \neq \mu_0$$

Level α test
Under H_0: $T = \sqrt{n}(\bar{X} - \mu_0) \sim N(0,1)$. Rejection region (based on T):

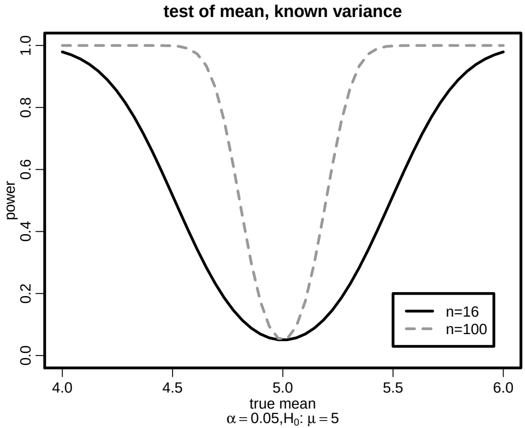
$$(-\infty, -c_{\alpha/2}] \cup [c_{\alpha/2}, \infty),$$

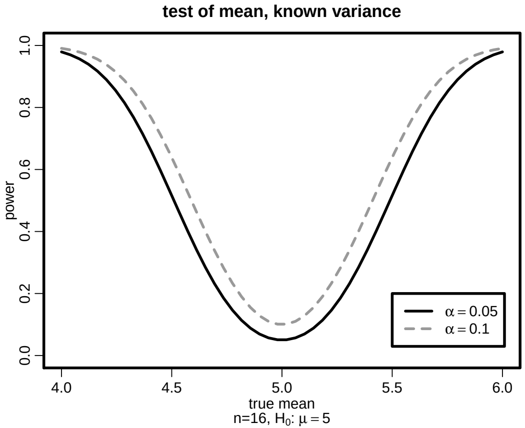
where $\Phi(c_{\alpha/2}) = 1 - \alpha/2$.
Test rejects for large values of $|T|$.
Hence, for the observation t the p-value is:

$$p = P_{\mu_0}(|T| \geq |t|) = P(T \leq -|t| \text{ or } T \geq |t|) = \Phi(-|t|) + 1 - \Phi(|t|) = 2 - 2\Phi(|t|)$$

Power: Note that $T \sim N(\sqrt{n}(\mu - \mu_0), 1)$.

$$\begin{aligned}\beta(\mu) =& P_\mu(|T| \geq c_{\alpha/2}) = 1 - P_\mu(-c_{\alpha/2} \leq T \leq c_{\alpha/2}) \\ =& 1 - P_\mu(-\sqrt{n}(\mu - \mu_0) - c_{\alpha/2} \leq T - \sqrt{n}(\mu - \mu_0) \leq -\sqrt{n}(\mu - \mu_0) + c_{\alpha/2}) \\ =& 1 - \Phi(-\sqrt{n}(\mu_0 - \mu) + c_{\alpha/2}) + \Phi(-\sqrt{n}(\mu_0 - \mu) - c_{\alpha/2})\end{aligned}$$





Example: Student's t-Test; One-Sample t-Test

$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ iid, μ and σ unknown parameters

$$H_0 : \mu = \mu_0 \text{ against } H_1 : \mu \neq \mu_0$$

Under H_0: $T = \sqrt{n}\frac{\bar{X}-\mu_0}{S} \sim t_{n-1}$.
Rejection region:

$$(-\infty, -c] \cup [c, \infty),$$

where $c = t_{n-1,\alpha/2}$. ($t_{n-1,\alpha/2}$ is chosen such that if $Y \sim t_{n-1}$ then $P(Y > t_{n-1,\alpha/2}) = \alpha/2$)

Relating Tests and CIs

Constructing a test from a confidence region

Let Y be the random observations. Suppose $A(Y)$ is a $1-\alpha$ confidence region for θ, i.e.

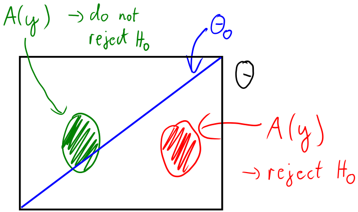
$$P_\theta(\theta \in A(Y)) \geq 1-\alpha \quad \forall \theta \in \Theta.$$

Then one can define a test for

$$H_0 : \theta \in \Theta_0 \quad \text{v.s.} \quad H_1 : \theta \notin \Theta_0$$

(where Θ_0 is some fixed subset of Θ) with level α as follows:

$$\text{Reject } H_0 \text{ if } \ \Theta_0 \cap A(y) = \emptyset.$$



To see that the above test has the appropriate level: For all $\theta \in \Theta_0$,

$$P_\theta(\text{type I error}) = P_\theta(\text{reject}) = P_\theta(\Theta_0 \cap A(Y) = \emptyset) \leq P_\theta(\theta \notin A(Y)) \leq \alpha.$$

Example: $Y \sim \text{Binomial}(n, \theta)$, $\theta \in (0, 1)$ unknown

$\sqrt{n}\frac{Y/n-\theta}{\sqrt{\theta(1-\theta)}}$ is approx. $N(0, 1)$

Leads to the confidence limits

$$\frac{y}{n} \pm \frac{c_{\alpha/2}}{\sqrt{n}}\sqrt{\frac{y}{n}(1-\frac{y}{n})}$$

A level α test of $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ is defined by

$$R = \left\{y : \theta_0 \notin \left(\frac{y}{n} - \frac{c_{\alpha/2}}{\sqrt{n}}\sqrt{\frac{y}{n}(1-\frac{y}{n})}, \frac{y}{n} + \frac{c_{\alpha/2}}{\sqrt{n}}\sqrt{\frac{y}{n}(1-\frac{y}{n})}\right)\right\}$$

Next lecture

We consider how to use parametric likelihoods to derive further tests