Hypothesis Testing

Dr. Romesh Thanuja

Outline

INTRODUCTION

DEFINITION

A hypothesis is a statement about a population parameter.

DEFINITION

The two complementary hypotheses in a hypothesis testing problem are called the null hypothesis and the alternative hypothesis. they are denoted by H_0 and H_1.

GENERAL HYPOTHESIS

For a vector of parameters θ and parameter space Θ, we test

$$H_0 : \theta \in \Theta_0 \text{ vs } H_1 : \theta \in \Theta_0^c$$

Note that Θ_0 and Θ_0^c must be mutually exclusive such that it is impossible for H_0 and H_1 to be true simultaneously.

Simple versus composite hypothesis

Simple versus composite hypothesis

A simple hypothesis takes the form

$$H_0 : \theta = c$$

where c is a constant, i.e. it states an exact specification of θ.

If a vector of parameters is being tested, then we may write $H_0 : \theta = c$, where c is a vector of constants of the same dimension as θ.

A hypothesis which is not simple is called a composite hypothesis, for example

$$H_0 : \theta \leq c.$$

Test Statistics

- The crucial part of a hypothesis test is to identify an appropriate test statistic, $T(\mathbf{X})$.
- One of the desired features of $T(\mathbf{X})$ is to have an interpretation such that large values of it provide evidence against H_0.
- We also want to know the distribution of $T(\mathbf{X})$ under H_0, that is when H_0 holds, we require the distribution of $T(\mathbf{X})|H_0$

Decision Rule

- Classical hypothesis testing involves a binary choice such that we: reject H_0 or not reject H_0
- Note a hypothesis testing procedure is incapable of "accepting" a hypothesis- a test can only check to what extent the available evidence is consistent or inconsistent with the null hypothesis.
- The evidence used to make the decision is given by the observed sample $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$. A test statistics, $T(\mathbf{X})$, is constructed and the computed test statistics value, $T(\mathbf{x})$, is used to make the decision of whether or not to reject H_0.
- The decision-making process can be summarized using a single function $\delta(x)$, as follows:

$$\delta(x) = \begin{cases} 1 & \text{when } H_0 \text{ is rejected} \\ 0 & \text{when } H_0 \text{ is not rejected} \end{cases}$$

The function δ is known as the decision rule. δ partitions the space of possible observed samples into two sets:

$$R = \{\mathbf{x} : \delta(\mathbf{x}) = 1\} \text{ and } R^c = \{\mathbf{x} : \delta(\mathbf{x}) = 0\}$$

where the set R is the rejection region (critical region) and R^c is the non-rejection region.

Error Probabilities

- A hypothesis test of $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_0^c$ might make one of two types of errors.
- If $\theta \in \Theta_0$, but the hypothesis test incorrectly decides to reject H_0, then the test has made a Type I error.
- If $\theta \in \Theta_1$ but the hypothesis test decides to accept H_0, a Type II error has been made.
- Suppose R denotes the rejection region for a test. Then for $\theta \in \Theta_0$, the test will make a mistake if $\mathbf{x} \in R$, so the probability of a Type I Error is $P_\theta(\mathbf{X} \in R)$
- For $\theta \in \Theta_0^c$, the probability of a Type II Error is $P(X \in R^c)$ (Note that $P_\theta(\mathbf{X} \in R^c) = 1 - P_\theta(\mathbf{X} \in R)$).
- We have

$$P_\theta(\mathbf{X} \in R) = \begin{cases} \text{probability of a Type I Error} & \text{if } \theta \in \Theta_0 \\ \text{one minus the probability of a Type II Error} & \text{if } \theta \in \Theta_0^c \end{cases}$$

POWER FUNCTION

DEFINITION

The power function of a hypothesis test with rejection region R is the function of θ defined by $\beta(\theta) = P_\theta(\mathbf{X} \in R)$

DEFINITION

For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a size α test if $\sup_{\theta \in \Theta_o} \beta(\theta) = \alpha$

DEFINITION

For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a level (significant) α test if $\sup_{\theta \in \Theta_o} \beta(\theta) \leq \alpha$

p-VALUE

p-VALUE

Let $T(X)$ be a test statistic used in a hypothesis testing procedure of a scalar parameter θ. Without loss of generality, suppose H_0 is rejected for large values of $T(\mathbf{X})$. For an observed sample $\mathbf{x}$ the corresponding p-value is

$$p(x) = \sup_{\theta \in \Theta_0} P(T(\mathbf{X}) \geq T(x))$$

where $T(\mathrm{x})$ is the test statistics value.

Note that: If we have a fixed significance level α, then we can describe the rejection region as:

$$R = \{\mathbf{x} : p(x) \leq \alpha\}$$