

Litterature review - Hypothesis testing

Département de mathématiques et de statistique - Université Laval

Stéphane Caron

08 December, 2019

Contents

1	Introduction	1
2	Simple hypothesis testing	1
2.1	p -value	2
2.2	Type of errors	2
3	Multiple comparison procedures	2
3.1	Family-wise error rate	2
3.2	False discovery rate	2
4	Examples of anomaly tests	2
4.1	Chi-square test	2
	Hoeffding's Universal Test	2

1 Introduction

This document provides a quick review of hypothesis testing theory. We first cover the basic of simple hypothesis testing. Then, we'll cover briefly some concepts about multiple hypothesis testing and also why it is relevant in the context of reliable anomaly detection. Finally, we'll describe 2 tests that might be used in order to conclude if an observation comes from a certain distribution (inlier) or not (outlier).

2 Simple hypothesis testing

A hypothesis test is a method of statistical inference. It aims to test a **null hypothesis** (H_0) versus an **alternative hypothesis** (H_1). In theory, the best way to test an hypothesis would be to gather data about the entire population and test our hypothesis. However, in real life we often do not have access to data of the entire population. In fact, we usually have access to a sample only, on which we'll use the framework of hypothesis testing to confirm or infirm a certain assumption.

As an example, let's say we want to test if the average height of the canadian male adults is greater or equal than 1.75 m or less. Indeed, we probably don't have access to the informations of all canadians, we then need to test on a given sample of canadians. We can formulate our test like that :

$$\begin{cases} H_0 : \mu \geq 1.75 \\ H_0 : \mu < 1.75 \end{cases}$$

where μ is the average height of canadian male adults.

Once we have our hypothesis, we need a test method that will includes statistical assumptions about our sample. In our case, we can use the one-sample t-test, that determines whether the hypothesized mean (μ) differs from the observed sample mean (\bar{X}). That will allows us to compute a certain statistic and then

compare it with a known distribution. This example is considered a simple hypothesis test since we do one single test. We'll cover some concepts of testing more than one hypothesis at the time in further sections.

2.1 p -value

Once we have a computed test statistic coming from a known distribution, we can compute what we call a p -value. A p -value is defined as the probability, under null hypothesis, to observe the test statistic from the known distribution. When we have a small the p -value, it means that the hypothesis may not adequately explain the observation. The null hypothesis is then rejected when value is less than a certain threshold α , which is referred to as the **level of significance**.

When the null hypothesis is true and the underlying random variable is continuous, then the probability distribution of the p -value is uniform on the interval $[0, 1]$.

2.2 Type of errors

When a decision is made from comparing a p -value with a level of significance, two kinds of errors can occur:

- Type I error (false positive): occurs when the null hypothesis is rejected when it is in fact true. The probability of making that error is called the **significance level** (also called α).
- Type II error (false negative): occurs when failing to reject null hypothesis that is false. The probability of committing that kind of error is often called **Beta**. The probability of not committing a type II error is called the **power** of the test.

3 Multiple comparison procedures

Multiple comparison procedures refers to the testing of more than one hypothesis at a time (Shaffer 1995). This framework aims to counteract the problem of multiple comparisons, where the more hypotheses we check, the higher the probability of Type I error (false positive) arises. The multiple comparison procedures often adjust the rejection criteria for each individual hypotheses, so that we do not have a single level of significance.

3.1 Family-wise error rate

3.2 False discovery rate

4 Examples of anomaly tests

4.1 Chi-square test

Hoeffding's Universal Test

Shaffer, Juliet. 1995. "Multiple Hypothesis Testing." *Annual Review of Psychology* 46 (January): 561–84. <https://doi.org/10.1146/annurev.ps.46.020195.003021>.