EX 05 - Intro to hypothesis tests

Afek Adler

2019-11-23

Last exercise we did:

- Talk on interval estimation
- Go throw a reminder on MLE with a motivation for machine learning (a mixture process)

Today we talk about:

- The null hypothesis
- General framework hypothesis testing
- Type 1 and type 2 errors
- P value
- The connection between hypothesis testing and confidence intervals
- Hypothesis tests

Some more topics that were covered in the lecture (will not be in the exercise):

- Type-II error and determining the sample size
- QQ plot (comparing distributions)

1 The null hypothesis

In inferential statistics, the null hypothesis is a general statement or default position that there is nothing new happening, like there is no association among groups, or no relationship between two measured phenomena. Testing (accepting, approving, rejecting, or disproving) the null hypothesis—and thus concluding that there are or are not grounds for believing that there is a relationship between two phenomena (e.g. that a potential treatment has a measurable effect)—is a central task in the modern practice of science; the field of statistics gives precise criteria for rejecting a null hypothesis. The null hypothesis is generally assumed to be true until evidence indicates otherwise. In statistics, it is often denoted H0, pronounced as "H-nought", "H-null", or "H-zero" (or, even, by some, "H-oh"), with the subscript being the digit 0.

The concept of a null hypothesis is used differently in two approaches to statistical inference. In the significance testing approach of Ronald Fisher, a null hypothesis is rejected if the observed data are significantly unlikely to have occurred if the null hypothesis were true. In this case, the null hypothesis is rejected and an alternative hypothesis is accepted in its place. If the data are consistent with the null hypothesis, then the null hypothesis is not rejected. In neither case is the null hypothesis or its alternative proven; the null hypothesis is tested with data and a decision is made based on how likely or unlikely the data are. This is analogous to the legal principle of presumption of innocence, in which a suspect or defendant is assumed to be innocent (null is not rejected) until proven guilty (null is rejected) beyond a reasonable doubt (to a statistically significant degree). https://en.wikipedia.org/wiki/Null_hypothesis (wiki)
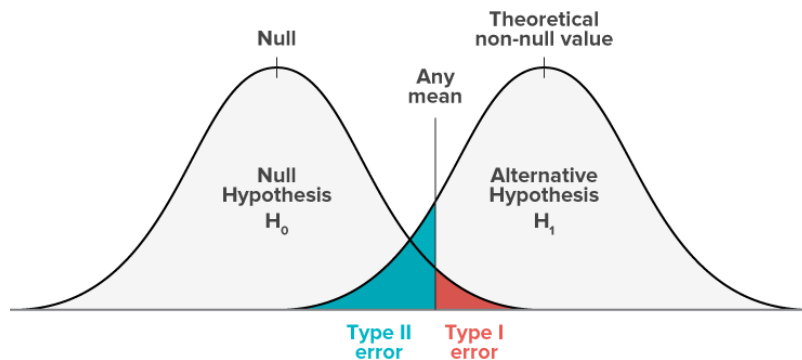
2 General Framework for hypothesis testing

This is the procedure for hypothesis testing:

1. Identify the parameter of interest (i.e., proportion, expectancy, std, etc.)

2. State the null hypothesis H_0
3. Specify the alternative hypothesis H_1 (one sided, two sided, etc.)
4. Choose significance level
5. Determine what test statistic to use (e.g., Z, T, X^2)
6. State the rejection region for the statistic
7. Compute the sample quantities, plug-in into the test statistic and compute it
8. Decide if should be rejected based on 6-7

3 Type 1 and Type 2 errors

On board:

- α - the probability if H_0 is True but our test says otherwise ()
- $1 - \alpha$
- β the probability if H_1 is True but our test says otherwise
- $1 - \beta$
- p_{value} - Intuition - low p_{value} means it's very likely that H_0 generated our sample
- Rejection region C and it's counterpart, the acceptance region $\overline{C}$

4 The relationship between hypothesis testing and confidence intervals

It can be shown that H_0 is accepted if and only the confidence interval contains the parameter in the basis of the assumption of H_0. For example, let's look at the hypothesis test for the mean:

Acceptance region -

$$\mu_0 - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu_0 + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

And by re-arranging both sides we get:

$$\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu_0 \leq \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

Which is exact the confidences intervals.

5 Hypothesis tests

5.1 Goodness of fit test:

Goodness of fit tests are used to test how good is the fit of our empirical distribution to that of a theoretical distribution. Arrange the empirical distribution in k bins, and let O_i be the observed frequency in the ith class bin. Let E_i be the expected probability.

H_0 : our observation are distributed according to ~Y (some distribution)
H_1 : else

The test statistic is:

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

We would reject the hypothesis if $\chi_0^2 > \chi_{\alpha,k-p-1}^2$

Where p is the number of parameters of the distribution.

Q1 - on board Q2 - on board