



Programming Assignment - Part 6

Submission Deadline:

12. Jun 2018

6 Statistic Tests (25 Points)

The last part of the programming assignment deals with statistic tests. As an example, you will implement the chi square test to check, if a set of samples fits an estimated distribution.

6.1 Implementation

The basic formulas of the chi square test have been presented in the lecture. You should implement them in the class *ChiSquare* in the file *statistictests.py*.

Task 6.1.1: Implementation of ChiSquare class

15 Points

Implement the functions *init()* and *test_distribution()*. The first function should take two input parameters:

- An array *emp_x* with $k+1$ values which represent the borders of the bins.
- An array *emp_n* with k values which represent the frequencies of each bin.

The second function should take the significance level α , mean value and standard deviation of a normal distribution as an input and should test the empirical values against the distribution and report, if the zero-hypothesis is rejected or not. Hint: Use the formulas provided in the lecture.

In addition, as is shown in the lecture slides, the minimum expected frequency of each interval is 5. If this is not the case, the combination of neighboring intervals should be done accordingly. You should consider the simplest case, in which only 100 samples are provided. Since the expected distribution is a normal distribution, only the intervals on both tails of the distribution should be combined as 2 compound intervals.

6.2 Verification

You should now verify your implementation of the test. You can simply do this by drawing samples from a normal distribution with given mean and variance. Then, you run your test on an distribution with the same parameters. Depending on the number of bins, the hypothesis will not be rejected in most cases. Do the coding for this section in file *part6_simstudy.py*.

In addition, you can run the given test *part6_tests.py* for verification. Note that for the test to work correctly, your *test_distribution()* must return the result: $[chi2, chi2_table]$, where *chi2* represents your calculated chi square value and the second variable represents the value drawn from the table (for given alpha and d.o.f.).

Task 6.2.1: Chi Square Test

5 Points

Draw 100 random samples from a standard normal distribution. Choose a reasonable number of bins and run the chi square test on your samples. The null hypothesis should be: the samples follow the standard normal distribution.

6.3 Analysis and General Questions

Answer the following questions separately and in full sentences. Explain your answers.

Task 6.3.1: Chi Square Test

5 Points

What can you observe, when you vary the number of bins or the significance level alpha in task 6.2.1? What happens, if you change the mean and/or the variance of the samples but still compare them to the standard normal distribution in the test? Describe your observations!

Total: 25 Points