Nonparametric Methods in Statistics

Bartłomiej Jamiołkowski

Testing the power of the Kolmogorov and Kolmogorov tests with the use of Probability Integral Transform in the case when the data come from the Student's t-distribution or a quadrilateral with a different number of data and degrees of freedom

1. Introduction

1.1 Project goal

The subject of the project is to investigate the power of the Kolmogorov and Kolmogorov tests using the PIT, with the main hypothesis that the feature has a normal distribution when the data come from the Student's t-distribution or the chi-square distribution. In the analysis, the test powers are examined taking into account:

- changes in the number of data;
- changes in the number of degrees of freedom.

1.2 Kolmogorov test

Before testing the power of the test, check whether its assumptions are met. A prerequisite for the use of the Kolmogorov test is the fact that this test is intended only for continuous distributions. In the project, the data is generated from Student's t-and chi-square distributions that satisfy this assumption.

In the conducted analysis, the null hypothesis of the test is the statement that the feature has a normal distribution represented by the distribution function $F0$ (x).

$$H0 : F(x) = F_0 (x)$$

$$H1 : F(x) \neq F_0 (x)$$

1.3 Kolmogorov test with the use of PIT

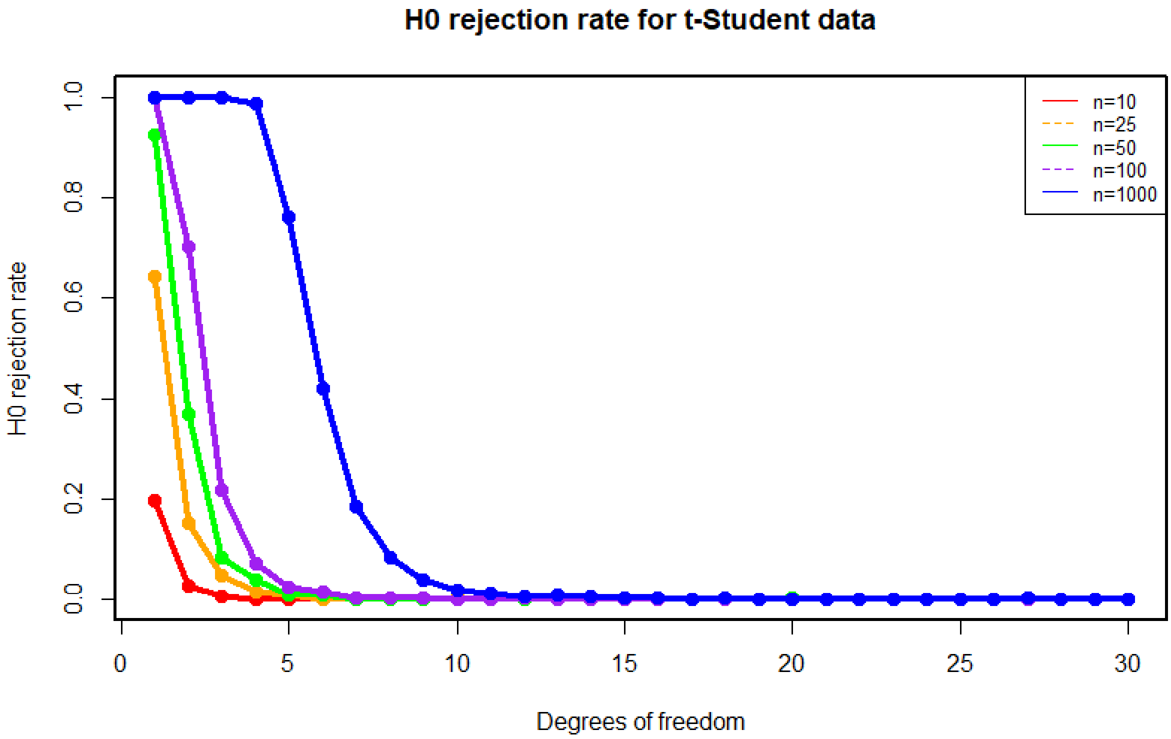
In the PIT (Probability Integral Transform) project, it is used in the Kolmogorov test to check whether the cumulative distribution of the generated distribution is continuous over the range 0-1 for the main hypothesis with a normally distributed trait (cumulative distribution $F0$ (x)).

1.4 Introduction to research

The study is divided into two parts. The first fragment examines the power of the Kolmogorov test, while the second examines the power of the Kolmogorov test with the use of PIT. In each of these parts, for both distributions, we check how the power of the test is affected by the change in the number of data in the sample and the degrees of freedom. The project uses samples with the number of data: 10, 25, 50, 100, and 1000. The degrees of freedom are in the range of integers from 1 to 30. The significance level of 0.05 is used in all tests. Based on the data from specific distributions and the calculated rejection rate H0, test power curves are generated in the RStudio program, on the basis of which conclusions are drawn.

2. Testing the power of the Kolmogorov test

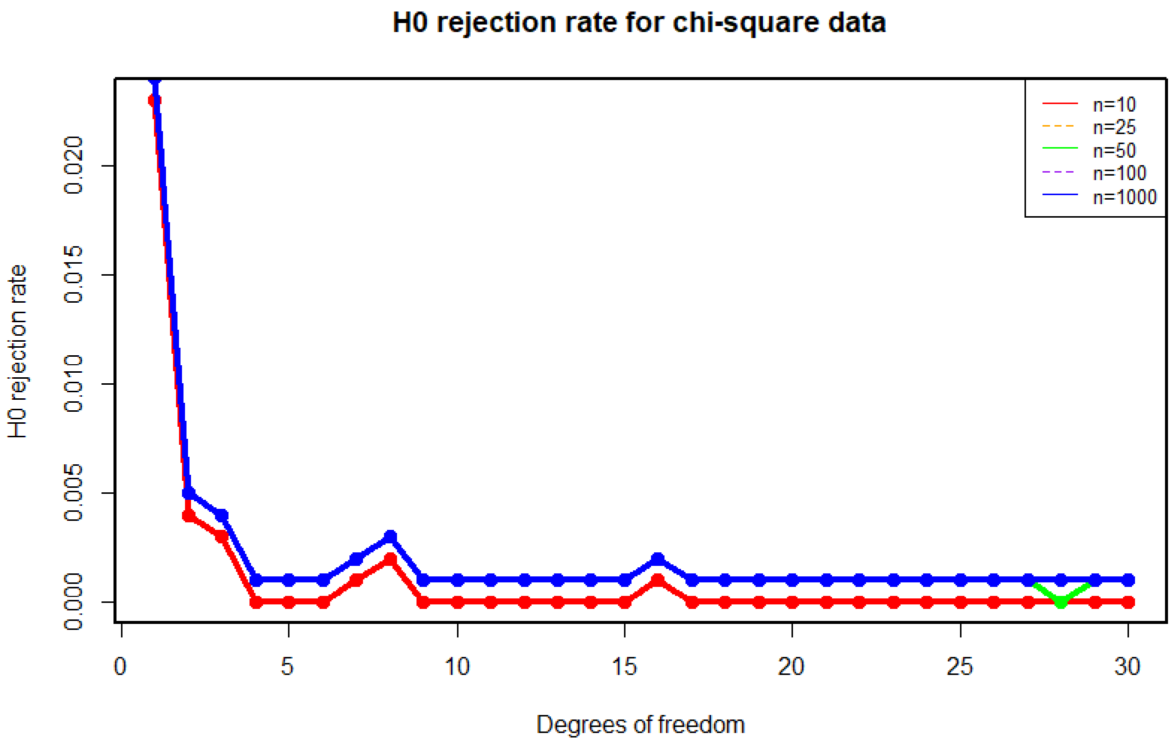
2.1 Testing the power of the Kolmogorov test for data from the Student's t distribution



Conclusions:

It can be seen in the presented graph that the power of the Kolmogorov test increases with the increase in the number of data in the samples. Level 1.0 rejects the null hypothesis, which is false when samples are generated from the Student t-distribution at 100%. At the same time, the graph shows that with the increase in the degrees of freedom, the H0 rejection rate decreases. For samples with the amount of data: 10, 25, 50 and 100, the decrease in test power is noticeable after 1 degree of freedom. For the sample with the number of data equal to 1000, the decrease is visible after the 3rd degree of freedom. From the 11th degree of freedom, the power of the test is close to 0.0 regardless of the sample size. The test shows high power for n = 1000, n = 100 and n = 50, with only samples with a data number of 1000 and 100 reaching 100% H0 rejection rates at the beginning of the measurements.

2.2 Investigation of the power of the Kolmogorov test for data from the chi-square distribution

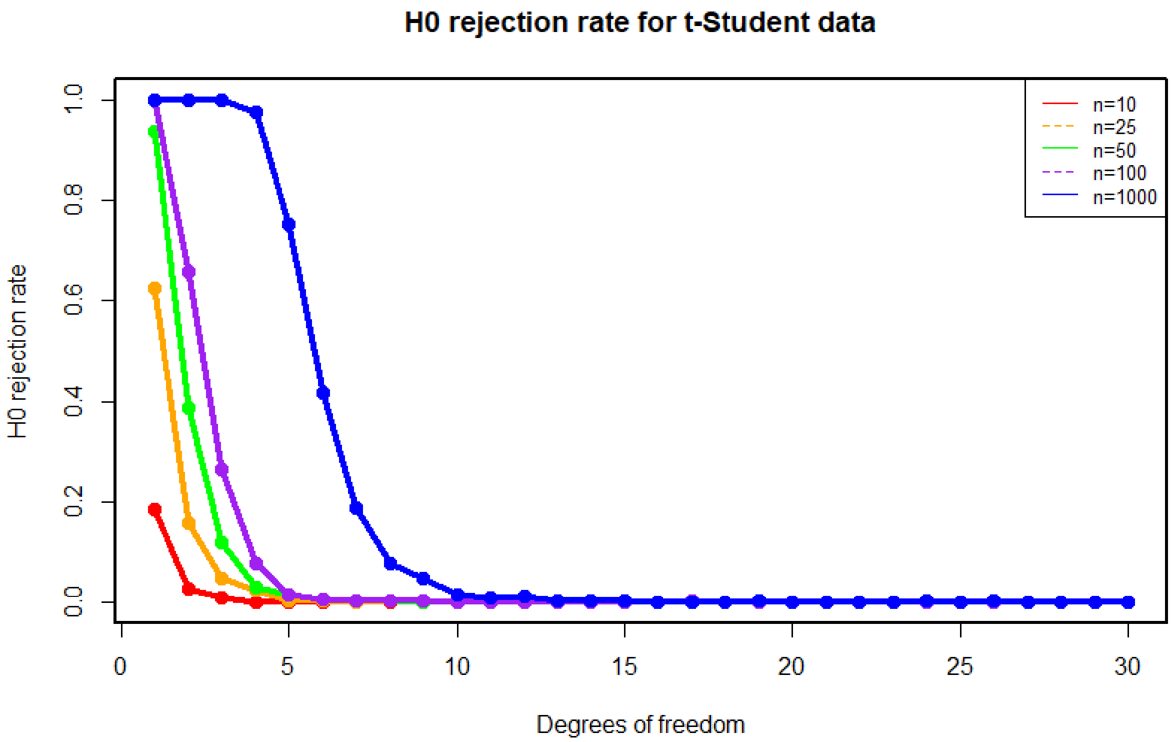


Conclusions:

The presented power plot of the Kolmogorov test for the data from the chi-square distribution has a different scale, because the reported test powers are very small. The highest percentage of rejections of the H0 hypothesis is repeated around the level of 0.025 and occurs for all samples with a different number of data with 1 degree of freedom. The graph shows that as the degrees of freedom increase, the power of the test decreases. These changes in the 0.0-1.0 test power scale are small. From the 5th degree of freedom, the power of the test reaches a level close to 0.0. For all samples in the plot at degrees of freedom 11, 15, and 16, this value is slightly larger. For a sample with a data number of 50 and a degree of freedom of 25, there is some deviation, and then the test power becomes 0.0. The graph and simulations in RStudio show that for samples with n = 10, the rejection rate H0 is slightly lower than for samples with the number of data: n = 25, n = 50, n = 100 and n = 1000.

3. Testing the power of the Kolmogorov test with the use of PIT

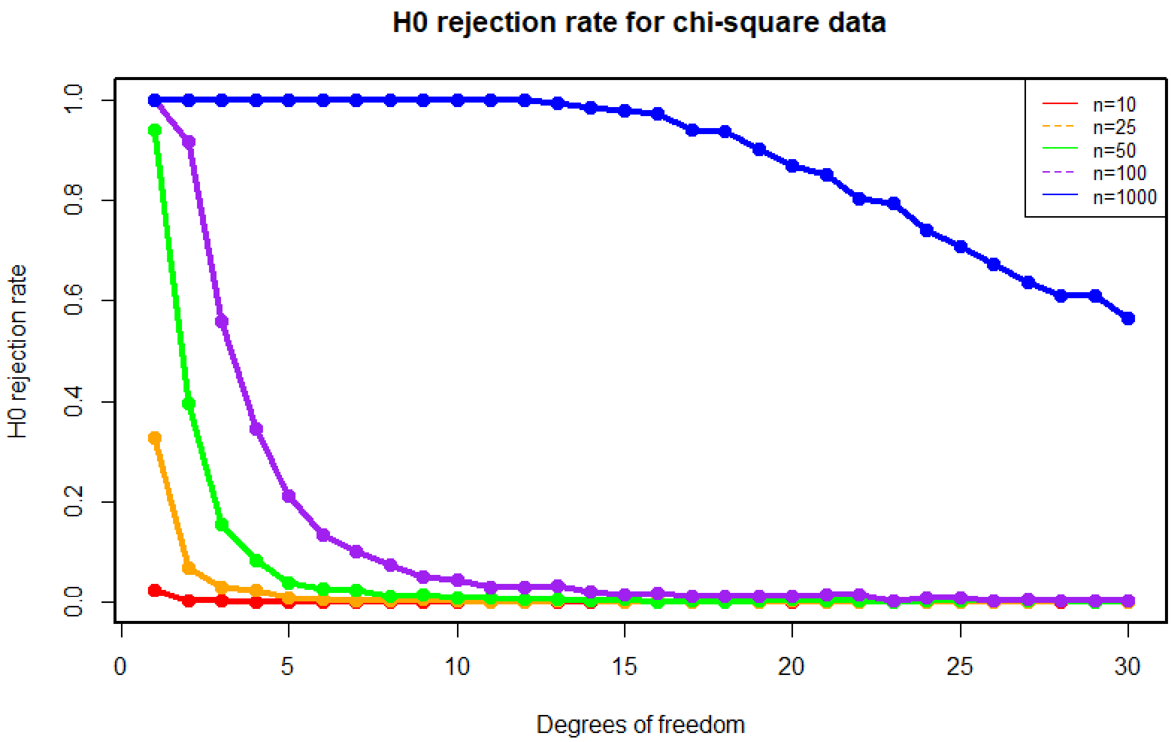
3.1 Testing the power of the Kolmogorov test with the use of PIT for data from the Student's t-distribution



Conclusions:

In the project, the Kolmogorov test with the use of PIT is used to test whether the data generated from the Student's t-distribution has a uniform distribution on the range 0 - 1. Based on the generated data, the values of the normal distribution function are calculated. The presented graph shows that with the increase in the number of data in the sample, the power of the test increases, while the increase in the degrees of freedom causes a decrease in the H0 rejection rate. From the 13th degree of freedom, regardless of the data size, the test power reaches a level close to 0.0. For data with: n = 50, n = 100 and n = 1000, the test power is very high, while only samples with a data size of 100 or 1000 achieve the test power of 100%. The graph also shows the values of the degrees of freedom, when the H0 bounce rate decreases. For samples with: n = 10, n = 25, n = 50 and n = 100, the power drop is visible after 1 degree of freedom. In the case of data with a cardinality of 1000, the decline occurs only after the 3rd degree of freedom.

3.2 Testing the power of the Kolmogorov test with the use of PIT for data from the chi-square distribution



Conclusions:

The presented diagram shows the powers of the Kolmogorov test using the PIT, examining whether the data generated from the chi-square distribution has a uniform distribution on the 0-1 range. The chart suggests that as the amount of data increases, the power of the test increases. The highest values are achieved for n = 50, n = 100 and n = 1000, while only for data with n = 100 or n = 1000, the H0 rejection level at the beginning of the measurements is 100%. For data with the length: 10, 25, 50 and 100, the decrease in the test power is noticeable after 1 degree of freedom, while for data with n = 1000, the decrease is observed after the 11th degree of freedom.

4. Summary

The conducted analysis shows that the lengths of the samples and the degrees of freedom of the distributions t-
Student and chi-square significantly affect the power of the Kolmogorov test. The conclusions are as follows:

1) An increase in the number of observations / sample length increases the power of the Kolmogorov test;
2) An increase in the number of degrees of freedom causes a decrease in the power of the Kolmogorov test.