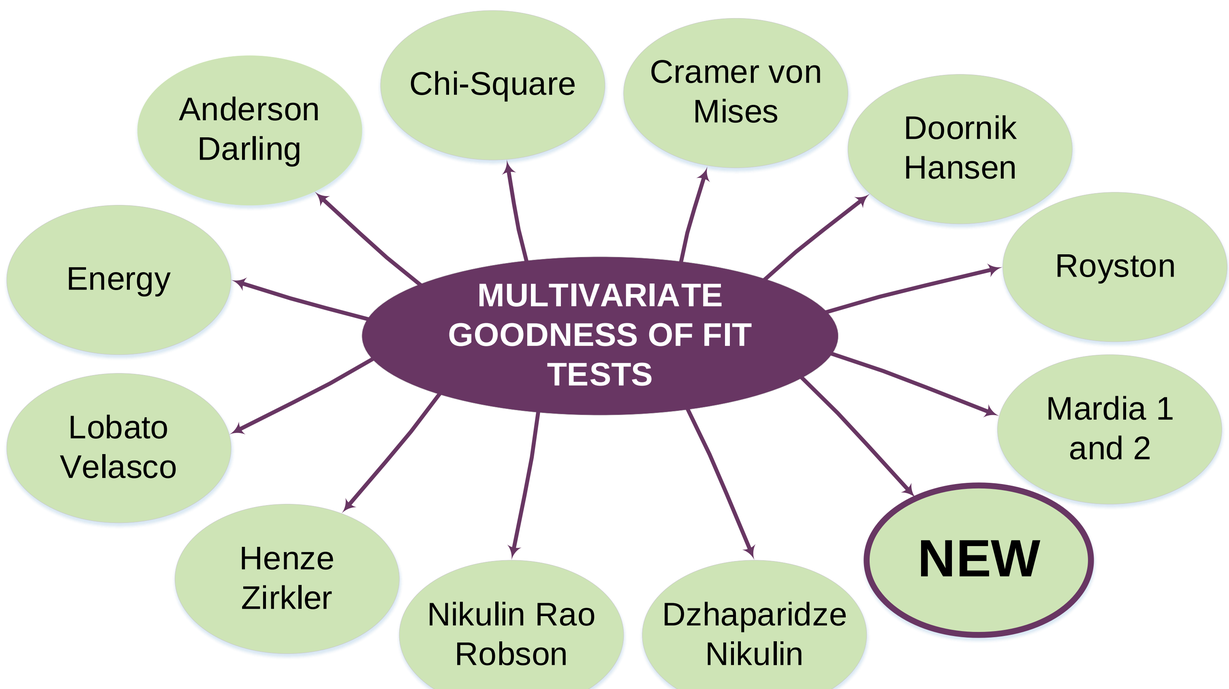
POWER ANALYSIS OF MULTIVARIATE GOODNESS OF FIT TESTS

kaunas university of technology
ktu 1922

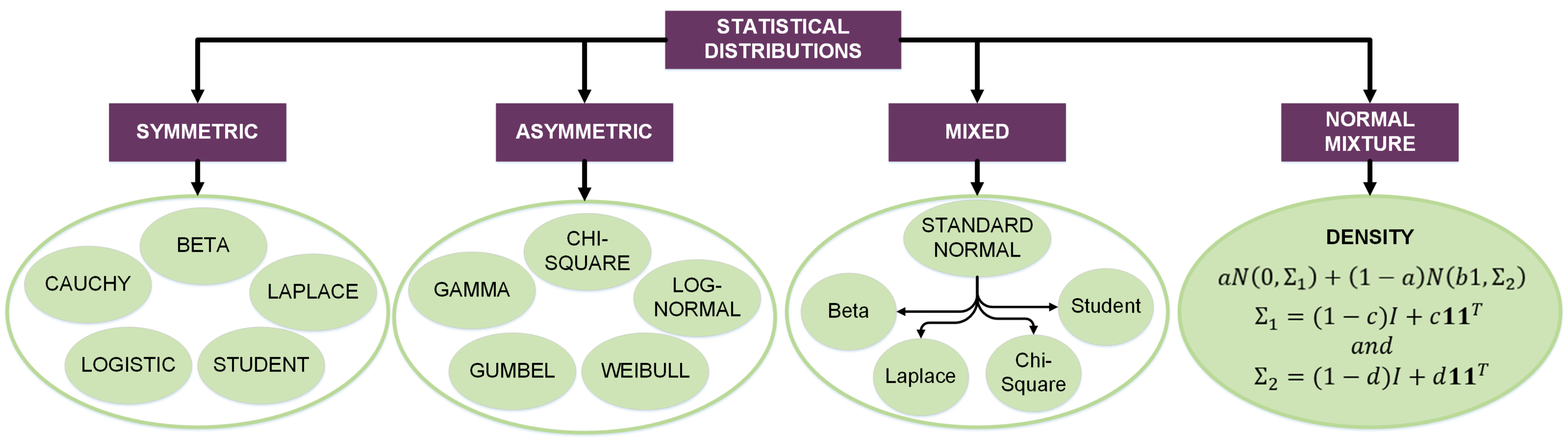
ABSTRACT

In modern data analytics, decisions making involves hypotheses testing. It is a common practice to check the assumption of data normality. Which dictates the choice of data analysis methods (parametric or non-parametric). The assumption of normality can be checked graphically, but a more consistent option is to test the goodness of fit hypothesis. Despite the fact that a lot of statistical test have been developed since the 20th century, analysis of multivariate data remains challenging. The purpose of this study is to perform a power analysis of multivariate goodness of fit hypothesis test for the assumption of normality for different data sets and to compare the results obtained with our proposed test. Our test is based on the mean absolute deviation of the empirical distribution density from the theoretical distribution density. In this test, the density estimate is derived by using a inversion formula. To show advantages of our test an exhaustive comparative study of multivariate tests was performed. For this purpose, a lot of multivariate data sets of non-normal distributions were generated. The power of well-known test and our test was evaluated empirically. Based on the obtained modelling results, it can be concluded that our test is more powerful than others in most of the cases.

MULTIVARIATE TESTS FOR NORMALITY



STATISTICAL DISTRIBUTIONS



SIMULATION STUDY

The study was performed for:

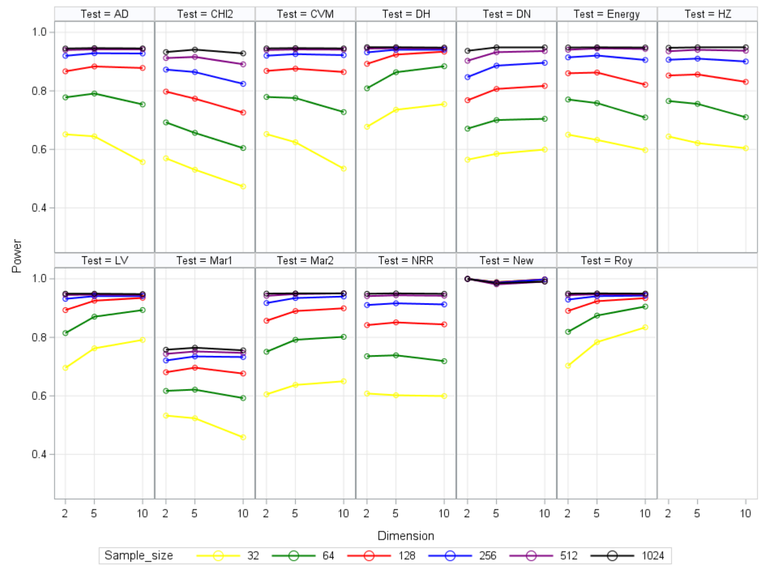
- **4** groups of distributions (i.e., **49** different data sets);
- **6** sample sizes (**32, 64, 128, 256, 512, 1024**);
- **12** multivariate goodness of fit tests;
- **3** dimensions (**2, 5, 10**);
- **1 000 000** independently generated samples.

CONCLUSIONS

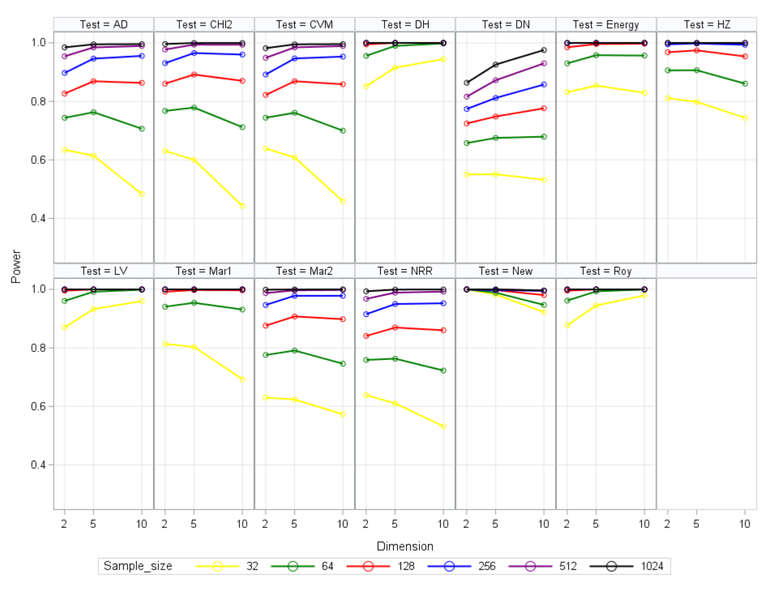
In this study, the comprehensive comparison of the power of 13 multivariate goodness of fit tests was performed for groups of symmetric, asymmetric, mixed, and normal mixture distributions. Two-dimensional, five-dimensional, and ten-dimensional data sets were generated to estimate the test power empirically.

A new multivariate goodness of fit test based on inversion formula was proposed. Based on the obtained modeling results, it was determined that the most powerful tests for the groups of symmetric, asymmetric mixed and normal mixture distributions are the proposed test and Roy multivariate test.

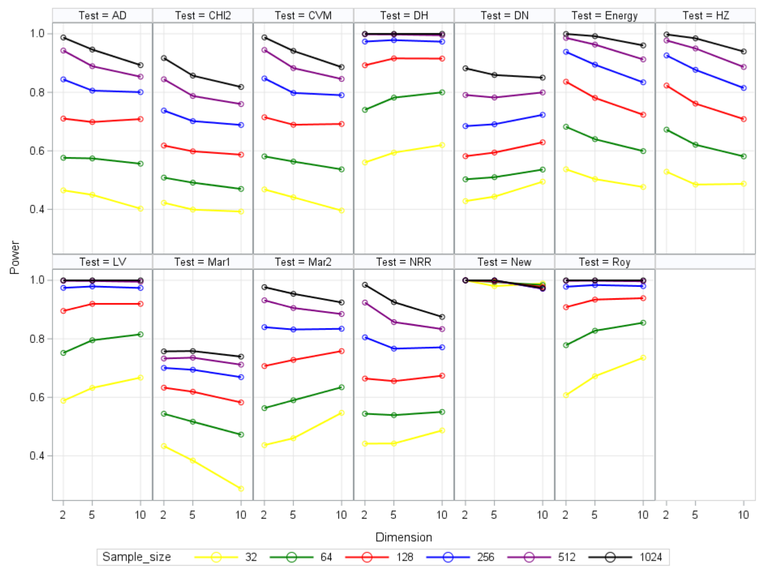
SYMETRIC GROUP



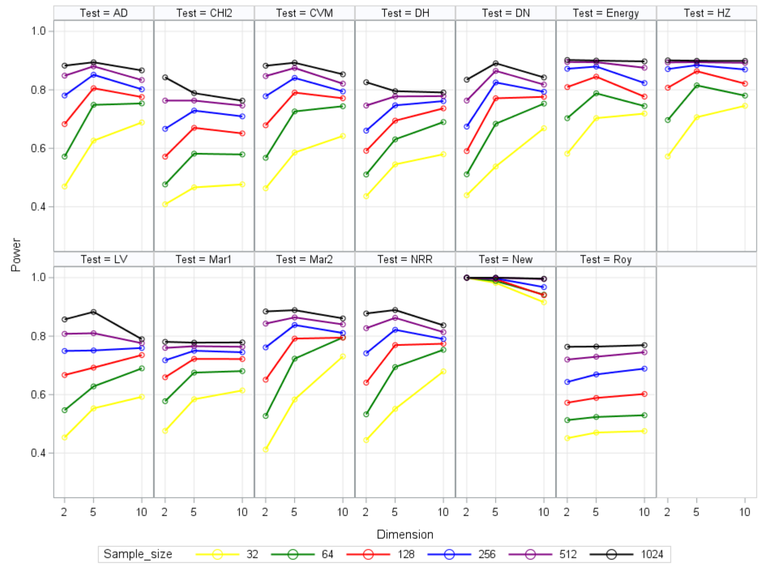
ASYMETRIC GROUP



MIXED GROUP



NORMAL MIXTURE GROUP



J. Arnastauskaitė,T. Ruzgas and M. Bražėnas