Exact Distribution-Free k-Sample Tests

for High Dimension, Low Sample Size Data

Shyamal Krishna De

Indian Statistical Institute, Kolkata, India

Abstract

In this talk, we consider the problem of testing equality of k (≥ 2) multivariate distributions when the dimension of the data is much larger than the sample size. Most of the existing tests often perform poorly in high dimension, low sample size (HDLSS) regime. We shall discuss some nonparametric tests that have the distribution-free property in finite sample situations. These tests are based on a high dimensional clustering algorithm that makes a partition of the data and produces a contingency table from which desirable test statistics are constructed. We consider tests based on a k-partition of the data as well as tests that uses number of partitions estimated from the data. Under appropriate assumptions, we prove consistency of these tests as dimension of the data grows to infinity when sample size remains fixed. We also discuss a multi-scale approach where the results for different number of partitions are aggregated judiciously. The superiority of the proposed tests over some exiting methods is shown by extensive simulation study. This is a joint work with Dr. Biplab Paul (University of Haifa, Israel) and Dr. Anil K. Ghosh (Indian Statistical Institute Kolkata, India).

Keywords:

Cluster Analysis; Contingency Tables; High Dimensional Asymptotics; Multi-Scale Approach; Tests of Independence.