

A Nonparametric Two-Sample Test Applicable to High Dimensional Data

Munmun Biswas and Anil K. Ghosh*

Theoretical Statistic and Mathematics Unit, Indian Statistical Institute,
203, B.T. Road, Kolkata 700108, INDIA

Email: munmun.biswas08@gmail.com, akghosh@isical.ac.in

Multivariate two-sample testing problem has been well investigated in the literature, and several parametric and nonparametric methods are available for it. However, most of these two-sample tests perform poorly for high dimensional data, and many of them are not applicable when the dimension of the data exceeds the sample size. In this article, we propose a nonparametric two-sample test that can be conveniently used in the high dimension low sample size set up. Asymptotic results on the power properties of our proposed test are derived when the sample size remains fixed, and the dimension of the data grows to infinity. We investigate the performance of this test on several high dimensional simulated and real data sets and demonstrate its superiority over several other existing two-sample tests. We also study the theoretical properties of the proposed test for situations when the dimension of the data remains fixed and the sample size tends to infinity. In such cases, it turns out to be asymptotically distribution-free and consistent under general alternatives.

Key Words: Inter-point distances, permutation tests, U-statistics, weak law of large numbers.