Statistica Sinica **36** (2026), 1-21
doi:https://doi.org/10.5705/ss.202023.0299

TESTING FOR THE EQUALITY OF DISTRIBUTIONS IN HIGH DIMENSION

Xu Li[1,2], Gongming Shi[1] and Baoxue Zhang*[1]

[1]*Capital University of Economics and Business and*
[2]*Shanxi Normal University*

Abstract: In this paper, we propose a new homogeneous test for two high-dimensional random vectors. Our test is built on a new measure, the so-called characteristic distance, which can completely characterize the homogeneity of two distributions. The newly proposed metric has some desirable properties, for example, it possesses a clear and intuitive probabilistic interpretation, and can be used to address the high-dimensional distance inference. Theoretically, the limiting behaviors under the conventional fixed dimension and high-dimensional distance inference are thoroughly investigated. Simulation studies and real data analysis are presented to illustrate the finite-sample performance of the proposed test statistic.

Key words and phrases: Characteristic distance, high dimensionality, permutation procedure, test of homogeneity, U-statistic.

1. Introduction

Over the past decades, the problem of assessing the homogeneity of two high-dimensional data has often appeared in various research areas. In some specific situations, the researchers want to measure whether two samples are generated from the same population. One example can be found in clustering analysis, where before constructing the groups, it is recommended to verify whether it is really necessary. For this, a formal test of the null hypothesis that two samples have been drawn from the single population is essential to prevent misjudgment.

The research on measuring and testing the homogeneity of two populations has a long history. For univariate data, the most traditional tools are the Smirnov maximum deviation test (Smirnoff, 1939) and Wald Wolfowitz runs (Wald and Wolfowitz, 1940), whose multivariate and multidimensional extensions have been widely discussed; examples include the Darling (1957), Bickel (1969), and Friedman and Rafsky (1979), among others. Recent years see another attempt to address the homogeneity between two random vectors by using the empirical characteristic function. Fernández, Gamero and García (2008) based on the empirical characteristic function proposed a class of tests for the two sample problems. Liu, Xia and Zhou (2015) and Liu, Liu and Zhou (2019) exploited jackknife empirical likelihood with empirical characteristic function to study the

*Corresponding author. E-mail: zhangbaoxue@cueb.edu.cn