Statistica Sinica **35** (2025), 1-21
doi:https://doi.org/10.5705/ss.202022.0366

A KERNEL INDEPENDENCE TEST USING PROJECTION-BASED MEASURE IN HIGH-DIMENSION

Yuexin Chen and Wangli Xu*

Renmin University of China

Abstract: Testing the independence between two high-dimensional random vectors is a fundamental and challenging problem in statistics. Most existing tests based on distance and kernel may fail to detect the non-linear dependence in the high-dimensional regime. To tackle this obstacle, this paper proposes a kernel independence test for assessing the independence between two random vectors based on a class of Gaussian projections relying on tuning parameters. The proposed test can be generally implemented for a wide class of distance-based kernels and completely characterizes dependence in the low-dimensional regime. Besides, the test captures pure non-linear dependence in the high-dimensional regime. Theoretically, we develop central limit theorem and associated rate of convergence for the proposed statistic under some mild regularity conditions and the null hypothesis. Moreover, we derive the asymptotic power of the proposed test enabling us to select suitable parameters for a special alternative, to achieve superior power in the high-dimensional regime. The choices of tuning parameters ensure that the proposed test has comparable power with the original kernel-based test in the moderately high-dimensional regime. Numerical experiments also demonstrate the satisfactory empirical performance of the proposed test in various scenarios.

Key words and phrases: High-dimension, independence test, kernel independence measure, random projections, U-statistics.

1. Introduction

Testing the independence of a pair of potentially high-dimensional random vectors has gained importance due to the increasing attention from big data applications (Kong, Wang and Wahba, 2015; Chakraborty and Zhang, 2019). Let $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ be two random vectors with probability measures P_X and P_Y, respectively. Given independent samples $\{(X_i, Y_i), i = 1, \ldots, n\}$ from P_{XY}, the hypothesis of interest is

$$\mathrm{H}_0 : P_{XY} = P_X P_Y \quad \mathrm{v.s.} \quad \mathrm{H}_1 : P_{XY} \neq P_X P_Y. \tag{1.1}$$

There exists a wide spectrum of dependency measures and tests. Notable examples include Pearson correlation (Pearson, 1895), rank correlation coefficients

*Corresponding author. E-mail: wlxu@ruc.edu.cn