Statistica Sinica **35** (2025), 1-21
doi:https://doi.org/10.5705/ss.202022.0388

CROSS PROJECTION TEST FOR HIGH-DIMENSIONAL MEAN VECTORS

Guanpeng Wang[1,2] and Hengjian Cui[*2]

[1] *Weifang University and* [2] *Capital Normal University*

Abstract: A cross projection test (CPT) technique for a one-sample vector in a high-dimensional setting is introduced. To overcome the problems caused by the curse of dimensionality, we construct test statistics by employing a projection test to project high-dimensional samples into one-or multi-dimensional directions. First, we randomly split the sample into two groups. We then find the p projection directions from a sample covariance matrix of the first group of samples. The second group is used to construct a projection statistic and perform the test. Second, we find the projection directions by exchanging the order of the two groups of samples, and we perform the test again to obtain another test statistic. Finally, we construct the CPT statistic by adding the two asymptotically uncorrelated test statistics together using the cross projection technique, such that the information from the two independent split samples can be fully utilized. The simulation results show that our proposed cross projection test controls the type I error well, and it is more powerful than the existing mean tests for some covariance matrix structures. Meanwhile, after applying the power enhancement technique, the CPT method performs non-trivially in general cases, especially for testing against sparse alternatives. A real gene-data analysis illustrates that the performance of our CPT is quite well.

Key words and phrases: Asymptotic distribution, cross projection test, high dimension, mean test, projection direction.

1. Introduction

It is well known that the hypothesis test of the mean vector is fundamental to multivariate statistical analysis (Anderson, 2003; Muirhead, 1982), which in turn is instrumental in diverse fields of research and application domains. The rapid development of technology has introduced new types of data, such as internet portals, hyperspectral imagery, microarray analysis, and DNA, to many fields. Generally speaking, these are often high-dimensional data in which the dimensionality of variables p is much larger than the sample size (n). This brings about the "curse of dimensionality" in statistical data analysis, which renders classical test statistics invalid or no longer applicable. The past two decades have witnessed increasing interest in mean signals difference identification for high-dimensional settings, and the existing methods are generally classified into

*Corresponding author. E-mail: hjcui@bnu.edu.cn