Statistica Sinica **37** (2027), 1-22
doi:https://doi.org/10.5705/ss.202024.0051

SPATIAL-SIGN BASED MAXSUM TEST FOR HIGH DIMENSIONAL LOCATION PARAMETERS

Jixuan Liu[1], Long Feng[*1], Ping Zhao[*2] and Zhaojun Wang[1]

[1]*Nankai University and* [2]*Tiangong University*

Abstract: In this study, we explore a robust testing procedure for the high-dimensional location parameters testing problem. Initially, we introduce a spatial-sign based max-type test statistic, which exhibits excellent performance for sparse alternatives. Subsequently, we demonstrate the asymptotic independence between this max-type test statistic and the spatial-sign based sum-type test statistic (Feng and Sun, 2016). Building on this, we propose a spatial-sign based max-sum type testing procedure, which shows remarkable performance under varying signal sparsity. Our simulation studies underscore the superior performance of the procedures we propose.

Key words and phrases: Asymptotic independence, high dimensional data, scalar-invariant, spatial-sign.

1. Introduction

The testing of location parameters is a crucial and extensively researched area in multivariate statistics with a fixed dimension. The conventional Hotelling's T^2 test is commonly applied, but it fails in high-dimensional scenarios where the variable's dimension p exceeds the sample sizes n. Consequently, numerous efforts have been made to develop a high-dimensional mean test procedure. One straightforward approach is to substitute the Mahalanobis distance with the Euclidean distance. For the two-sample location problem, Bai and Saranadasa (1996) employed the L_2-norm of the difference between two sample means. Chen and Qin (2010) eliminated some redundant terms in Bai and Saranadasa's (1996) test statistics and made no assumptions about the relationship between the dimension and sample sizes. Srivastava (2009), Park and Ayyala (2013), and Feng et al. (2015) suggested some scalar-invariant test statistics that replace the sample covariance matrix in Hotelling's T^2 test statistics with its diagonal matrix. All these methods are built on the assumption of normal distribution or diverging factor models, which perform poorly for heavy-tailed distributions. For instance, the well-known multivariate t-distribution does not meet the above assumption. Therefore, numerous studies have also considered robust high-dimensional test procedures.

*Corresponding author. E-mail: (Feng) flnankai@nankai.edu.cn; (Zhao) zpnanakai@hotmail.com