

# STATISTICAL INFERENCE FOR ULTRAHIGH DIMENSIONAL LOCATION PARAMETER BASED ON SPATIAL MEDIAN

Guanghui Cheng, Liuhua Peng and Changliang Zou\*

*Guangzhou University, The University of Melbourne and Nankai University*

*Abstract:* Motivated by the widely used geometric median-of-means estimator in machine learning, this paper studies statistical inference for ultrahigh dimensionality location parameter based on the sample spatial median under a general multivariate model, including simultaneous confidence intervals construction, global tests, and multiple testing with false discovery rate control. To achieve these goals, we derive a novel Bahadur representation of the sample spatial median with a maximum-norm bound on the remainder term, and establish Gaussian approximation for the sample spatial median over the class of hyperrectangles. In addition, a multiplier bootstrap algorithm is proposed to approximate the distribution of the sample spatial median. The approximations are valid when the dimension diverges at an exponentially rate of the sample size, which facilitates the application of the spatial median in the ultrahigh dimensional region. The proposed approaches are further illustrated by simulations and analysis of a genomic dataset from a microarray study.

*Key words and phrases:* Bootstrap approximation, FDR control, Gaussian approximation, high-dimensional, spatial median.

## 1. Introduction

Geometric median-of-means (GMOM) has been widely used for robust estimation of multivariate means, and it has been broadly adopted in machine learning (Minsker, 2015; Hsu and Sabato, 2016; Prasad et al., 2020). The idea of GMOM is to first divide the data into disjoint subsamples and calculate the empirical means of each of the subsamples. Then the GMOM estimator is computed as the spatial median (also called geometric median) of the obtained empirical means. The previous studies on the GMOM focused on establishing its non-asymptotic error bounds under certain heavy-tailed assumptions. Its distributional properties, which are essential for statistical inference, remain unknown.

High-dimensional data with the dimension increases to infinity as the number of observations goes to infinity have been encountered in many scientific disciplines. There is a growing evidence that the multivariate normal distribution

---

\*Corresponding author. E-mail: [nk.chlzou@gmail.com](mailto:nk.chlzou@gmail.com)