Statistica Sinica **32** (2022), 2007-2022
doi:https://doi.org/10.5705/ss.202020.0262

SPECTRAL PROPERTIES OF RESCALED SAMPLE CORRELATION MATRIX

Yanqing Yin, Changcheng Li, Guo-Liang Tian and Shurong Zheng

Chongqing University, Dalian University of Technology, Southern University of Science and Technology and Northeast Normal University

Abstract: Under the high-dimensional setting that the data dimension and sample size tend to infinity proportionally, we derive the limiting spectral distribution and establish the central limit theorem of the eigenvalue statistics of rescaled sample correlation matrices. In contrast to the existing literature, our proposed spectral properties do not require the Gaussian distribution assumption or the assumption that the population correlation matrix is equal to an identity matrix. The asymptotic mean and variance-covariance in our proposed central limit theorem can be expressed as one-dimensional or two-dimensional contour integrals on a unit circle centered at the origin. Not only is the established central limit theorem of the eigenvalue statistics of the rescaled sample correlation matrices very different to that of sample covariance matrices, it also differs from that of sample correlation matrices with a population correlation matrix equal to an identity matrix. Moreover, to illustrate the spectral properties, we propose three test statistics for the hypothesis testing problem of whether the population correlation matrix is equal to a given matrix. Furthermore, we conduct extensive simulation studies to investigate the performance of our proposed testing procedures.

Key words and phrases: Central limit theorem, limiting spectral distribution, random matrix theory, rescaled sample correlation matrix.

1. Introduction

With the rapid development of computer science, it is possible to collect, store, and analyze high-dimensional data sets. However, the classical statistical tools often are invalid when presented with such data. The high-dimensional sample correlation matrix is an important random matrix for principal component analysis, factor analysis, and human brain image analysis, among others. Let the sample $\mathbf{y}_1, \ldots, \mathbf{y}_n$ of size n be from a p-dimensional population $\mathbf{y}$ with unknown mean $\boldsymbol{\mu}$, covariance matrix $\boldsymbol{\Sigma}$, and correlation matrix

$$\mathbf{R} = [\mathrm{diag}(\boldsymbol{\Sigma})]^{-1/2}\boldsymbol{\Sigma}[\mathrm{diag}(\boldsymbol{\Sigma})]^{-1/2},$$

Corresponding author: Shurong Zheng, School of Mathematics and Statistics and KLAS, Northeast Normal University, Changchun, Jilin, China. E-mail: zhengsr@nenu.edu.cn.