Statistica Sinica **33** (2023), 1-20
doi:https://doi.org/10.5705/ss.202020.0327

EIGENVALUE DISTRIBUTION OF A HIGH-DIMENSIONAL DISTANCE COVARIANCE MATRIX WITH APPLICATION

Weiming Li, Qinwen Wang and Jianfeng Yao

Shanghai University of Finance and Economics,
Fudan University and The Chinese University of Hong Kong (Shenzhen)

Abstract: We introduce a new random matrix model called the distance covariance matrix, the normalized trace of which is equivalent to the distance covariance. We first derive a deterministic limit for the eigenvalue distribution of the distance covariance matrix when the dimensions of the vectors and the sample size tend to infinity simultaneously. This limit is valid when the vectors are independent or weakly dependent through a finite-rank perturbation. It is also universal and independent of the distributions of the vectors. Furthermore, the top eigenvalues of the distance covariance matrix are shown to obey an exact phase transition when the dependence of the vectors is of finite rank. This finding enables the construction of a new detector for weak dependence, where classical methods based on large sample covariance matrices or sample canonical correlations may fail in the considered high-dimensional framework.

Key words and phrases: Distance covariance, distance covariance matrix, eigenvalue distribution, finite-rank perturbation, nonlinear correlation, spiked models.

1. Introduction

Székely, Rizzo and Bakirov (2007) introduced the concept of the *distance covariance* $\mathcal{V}(\mathbf{x},\mathbf{y})$ of two random vectors $(\mathbf{x},\mathbf{y}) \in \mathbb{R}^p \times \mathbb{R}^q$ as a measure of their dependence. It is defined through an appropriately weighted L_2-distance between the joint characteristic function $\phi_{\mathbf{x},\mathbf{y}}(s,t)$ of $(\mathbf{x},\mathbf{y})$ and the product of their marginal characteristic functions $\phi_{\mathbf{x}}(s)\phi_{\mathbf{y}}(t)$, namely

$$\mathcal{V}(\mathbf{x},\mathbf{y}) = \left\{ \frac{1}{c_p c_q} \iint_{\mathbb{R}^p \times \mathbb{R}^q} \frac{|\phi_{\mathbf{x},\mathbf{y}}(s,t) - \phi_{\mathbf{x}}(s)\phi_{\mathbf{y}}(t)|^2}{\|s\|^{1+p}\|t\|^{1+q}} ds dt \right\}^{1/2}, \tag{1.1}$$

where the normalization constants are $c_d = \pi^{(1+d)/2}/\Gamma((1+d)/2)$ $(d = p, q)$. Clearly, $\mathcal{V}(\mathbf{x},\mathbf{y}) = 0$ if and only if $\mathbf{x}$ and $\mathbf{y}$ are independent.

For a collection of independent and identically distributed (i.i.d.) observa-

Corresponding author: Qinwen Wang, School of Data Science, Fudan University, Shanghai 200433, China. E-mail: wqw@fudan.edu.cn.