JIRSS (2021)
Vol. 20, No. 01, pp 197-218
DOI:10.52547/jirss.20.1.197

On the Blocks of Interpoint Distances

Reza Modarres [1]

[1] Department of Statistics, George Washington University, Washington, DC, USA.

Received: 30/11/2020, Revision received: 23/12/2020, Published online: 03/04/2021

Abstract. We study the blocks of interpoint distances, their distributions, correlations, independence and the homogeneity of their total variances. We discuss the exact and asymptotic distribution of the interpoint distances and their average under three models and provide connections between the correlation of interpoint distances with their vector correlation and test of sphericity. We discuss testing independence of the blocks based on the correlation of block interpoint distances. A homogeneity test of the total variances in each block and a simultaneous plot to visualize their relative ordering are presented.

Keywords. Elliptical Model, Sphericity, Homogeneity, Total Variance.

MSC: 62H10, 62H15, 62H20.

1 Introduction

Euclidean interpoint distances (IPDs) provide a versatile tool for analysis of high dimensional data. Interpoint distances are the building blocks of methods such as machine learning, classification, multidimensional scaling, discriminant analysis, depth functions, and testing homogeneity of distributions. Multivariate methods that depend on nonsingular covariance matrices fail to work in high dimensional settings

Corresponding Author: Reza Modarres (reza@gwu.edu)

when there are fewer number of observations n than variables p. Methods based on IPDs circumvent difficulties of high dimensional space because IPDs are always one dimensional.

IPDs are utilized in numerous applications, including geodesy, economics, genetics, psychology, biochemistry, and engineering. For example, Marozzi (2016) discusses multivariate tests based on IPDs and applies them to analyze magnetic resonance images. Guo and Modarres (2019) use IPDs for classification of high dimensional discrete observations and Song and Modarres (2020) use IPDs for testing homogeneity of multivariate mixture models. Modarres and Song (2020) survey recent developments on the Euclidean interpoint distances. Other distances than the Euclidean one have been used to develop powerful multivariate test for comparing locations (Marozzi 2015) and distributions (Marozzi et al. 2020), in particular when the underlying distributions are heavy-tailed or highly-skewed. Pal et al. (2016), Sarkar et al. (2020), and Modarres (2020) have proposed dissimilarity measures for high-dimensional, low sample size settings that use the differences of IPDs for clustering, testing the homogeneity of distributions, and outlier detection.

Let $\mathbf{X}_i = (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{ip})'$, $i = 1, \ldots, n$, be p dimensional i.i.d random vectors drawn from a population F with $\mathbb{E}(\mathbf{X}_i) = \boldsymbol{\mu}$, and covariance matrix $\mathbb{C}\mathrm{ov}(\mathbf{X}_\mathrm{i}) = \boldsymbol{\Sigma} > 0$. We partition $\mathbf{X}_i$ into k blocks $\mathbf{X}_i = (\mathbf{X}_i^{(1)}, \mathbf{X}_i^{(2)}, \ldots, \mathbf{X}_i^{(k)})'$ with $p_1, p_2, \ldots, p_k$ components where $\sum_{r=1}^{k} p_r = p$. The vector of means and the covariance matrix are partitioned similarly as $\boldsymbol{\mu} = (\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)}, \ldots, \boldsymbol{\mu}^{(k)})'$ and $\boldsymbol{\Sigma} = \left(\boldsymbol{\Sigma}_{rt}\right)$ for $r, t = 1, \ldots, k$. Let $\|\mathbf{X}\| = (\mathbf{X}'\mathbf{X})^{1/2}$ be the Euclidean norm of $\mathbf{X}$. The squared interpoint distance between $\mathbf{X}_i$ and $\mathbf{X}_j$ is given by $d^2\left(\mathbf{X}_i, \mathbf{X}_j\right) = \|\mathbf{X}_i - \mathbf{X}_j\|^2$ for $1 \leq i < j \leq n$. Two IPDs $d^2\left(\mathbf{X}_i, \mathbf{X}_j\right)$ and $d^2\left(\mathbf{X}_h, \mathbf{X}_k\right)$ are dependent if they have an subscript in common. With n observations, there are $m = n(n-1)/2$ IPDs and $m(m-1)/2$ pairs of the IPDs, among which $m(n-2)$ pairs of distances are dependent.

Let $d^2\left(\mathbf{X}_i^r, \mathbf{X}_j^r\right) = \|\mathbf{X}_i^{(r)} - \mathbf{X}_j^{(r)}\|^2$ denote the squared distance between the r-th components of $\mathbf{X}_i$ and $\mathbf{X}_j$. It follows that $d^2\left(\mathbf{X}_i, \mathbf{X}_j\right) = \sum_{r=1}^{k} d^2\left(\mathbf{X}_i^r, \mathbf{X}_j^r\right) = \sum_{r=1}^{k} \|\mathbf{X}_i^{(r)} - \mathbf{X}_j^{(r)}\|^2$. Since $d^2\left(\mathbf{X}_i, \mathbf{X}_j\right)$ are identically distributed, we use the random variable $d^2_{(X)}$ to denote the interpoint distance between $\mathbf{X}_i$ and $\mathbf{X}_j$. Similarly, we use the random variable $d^2_{(X^{(r)})}$ to denote the interpoint distance between the r-th components $d^2\left(\mathbf{X}_i^r, \mathbf{X}_j^r\right)$. We consider a data structure with m IPDs for the k-tuple $(d^2_{(X^{(1)})}, \ldots, d^2_{(X^{(k)})})$.

Study of the blocks of interpoint distances is important for testing independence of

the blocks. High dimensional datasets are commonplace in finance, robotics, network analysis, medical imaging, machine learning, microarray and DNA analysis where independence testing is a central issue as it facilitates dimension reduction. Classical tests of independence (Anderson, 2003) are based on the asymptotic theory that n approaches infinity while the dimension p is fixed where $p < n$. Therefore, it is not effective in settings where p approaches infinity. Methods based on the IPDs are effective for large p when $p > n$. We will show that under two models $\mathbf{X}^{(r)}$ and $\mathbf{X}^{(s)}$ are independent if and only if the interpoint distances $d^2_{(X^{(r)})}$ and $d^2_{(X^{(s)})}$ are independent for $r \neq s = 1, \ldots k$.

We are also interested in determining whether the means of the blocks of IPDs $\eta_r = \mathbb{E}(d^2_{(X^{(r)})})$, $r = 1, \ldots, k$, are the same across the k blocks. We test the null hypothesis

$$L_0 := \eta_1 = \eta_2 = \ldots = \eta_k, \tag{1.1}$$

against general alternative $L_a : \eta_r \neq \eta_t$ for at least one pair (r, t) where $r \neq t = 1, \ldots k$.

Finding outliers in high dimensional data sets is a difficult task since with increasing dimensions the observations become sparse and outliers become masked by multiple dimensions. Study of the blocks of interpoint distances is also important in outlier detection techniques that are distance-based. One may apply distance-based outlier detection techniques to blocks of the variables to determine if outliers exist in lower dimensions. Similar to canonical correlation analysis, one maybe interested in identifying variables that are highly correlated within their block, but uncorrelated with variables in other blocks. Study of the correlation structure of the blocks allows one to determine unequal subsets of the original variables that are uncorrelated or highly correlated. Bottesch et al. (2016) use blocks of variables to improve the lower bound on Euclidean distances and speed up k-means clustering. We will show that the correlation between any two blocks equals their vector correlation and relate the squared coefficient of variation of the IPDs to the test of sphericity under the multivariate normal distribution.

The rest of the article is organized as follows. The next Section discusses testing the independence of k blocks. The distributions of IPDs under the $\mathcal{M}$, *normal* and *elliptical* models are discussed in Section 3. In Section 4, we consider the average squared IPD and examine its properties. Asymptotic distribution of the IPDs for fixed n as p approaches infinity, and for fixed p as n tends to infinity are discussed in Section 5. Section 6 provides a connection between correlation of the IPDs and their vector correlation and the test of sphericity under normality. Section 7 presents a method for comparing total variances of the blocks. A simultaneous plot of the total variances is

provided to visualize their homogeneity. The last Section is devoted to summary and conclusions. The proofs appear in the Appendix.

2 Testing Block Independence

We are interested to test the null hypothesis that the joint distribution of the IPD blocks factors as the product of the distribution of IPD blocks,

$$H_0 : \mathbb{P}(d^2_{(X^{(1)})} \leq x_1, \ldots, d^2_{(X^{k})} \leq x_k) = \prod_{r=1}^{k} \mathbb{P}(d^2_{(X^{(r)})} \leq x_r) \text{ for all } (x_1, \ldots, x_k) \in \mathbb{R}^k. \tag{2.1}$$

The number of blocks k and the assignment of the variables to the blocks are determined by the investigator. The p_r variables in the r-th block maybe the result of variable selection techniques and can constitute variable clusters. The choices are influenced by the tendency of high-dimensional data to contain hubs that frequently occur in nearest-neighbor lists of other points (Flexer and Schnitzer, 2015). With the above perspective, many interesting hypotheses can be stated about the blocks of variables. In particular, under the normal and the power-normal distributions, we show that H_0 is equivalent to the independence of the IPD blocks, that is, for $r \neq t = 1, \ldots, k$, $G_0 : \mathbb{C}\text{ov}(d^2_{(X^{(r)})}, d^2_{(X^{t})}) = 0$. It is well known that G_0 is true if and only if the component covariances are zero under the *normal model*. That is, $\mathbf{\Sigma}_{rs} = \mathbf{0}$ for $r \neq s = 1, \ldots k$.

Consider the Box-Cox power transformation defined by $Y_j = (X_j^{\lambda_j} - 1)/\lambda_j$ where $\lambda_j \neq 0$ and $Y_j = \ln X_j$ when $\lambda_j = 0$ for each variable X_j, Y_j in $\mathbf{X} = (X_1, X_2, ..., X_p)$ and $\mathbf{Y} = (Y_1, Y_2, ..., Y_p)$, respectively. Suppose random vector $\mathbf{Y}$ has a p-variate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$. Let $Q = \left(\mathbf{Y}_j - \boldsymbol{\mu}\right)' \Sigma^{-1} \left(\mathbf{Y}_j - \boldsymbol{\mu}\right)$. Freeman and Modarres (2005, 2006) define a power-normal distribution with p.d.f. $f(\mathbf{X} \mid \boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{\Sigma}) = \frac{1}{K} \cdot \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \prod_{j=1}^{p} X_j^{\lambda_j - 1} \exp(-\frac{1}{2}Q)$ where K is a normalizing constant that depends on $\boldsymbol{\lambda}, \boldsymbol{\mu}$, and $\mathbf{\Sigma}$. The value of K is often assumed to be 1. Under multivariate power-normal distribution, the dependence between the variables is no longer linear. Let ρ_x and ρ_y denote the coefficient of correlations in the bivariate power-normal and bivariate normal scales, respectively. Freeman and Modarres (2006) $\rho_y = 0$ if and only if $\rho_x = 0$. Statistics based on interpoint distances not only detect linear dependence, but also find non-linear relationships that are induced on the power-normal scale.

Guo and Modarres (2020) show that under the *normal*, and power-normal model independence between two vectors is equivalent to the independence between their squared interpoint distances. The following theorem extends the result to k groups.

Theorem 2.1. *Let* $\mathbf{X}_i = (\mathbf{X}_i^{(1)}, \mathbf{X}_i^{(2)}, \ldots, \mathbf{X}_i^{(k)})'$ *for* $i = 1, 2$ *be i.i.d. copies of the random vector* $\mathbf{X}$ *with a joint multivariate normal or power-normal distribution with mean* $\boldsymbol{\mu}$*, covariance matrix* $\boldsymbol{\Sigma}$*. Consider the r-th and s-th component IPDs* $\|\mathbf{X}_1^{(r)} - \mathbf{X}_2^{(r)}\|^2$ *and* $\|\mathbf{X}_1^{(s)} - \mathbf{X}_2^{(s)}\|^2$*, respectively, for* $r \neq s = 1, \ldots, k$*. The vectors* $\mathbf{X}^{(r)}$ *and* $\mathbf{X}^{(s)}$ *are independent if and only if the interpoint distances* $d^2_{(X^{(r)})}$ *and* $d^2_{(X^{(s)})}$ *are independent.*

Theorem 2.1 allows one to test the independence of k blocks of interpoint distances through a $k \times k$ matrix of their correlations. To test whether the r-th and s-th blocks are independent one simply computes the correlation coefficient between the m interpoint distances. Permutation testing allows one to test for significance of this correlation. Similarly, to determine if all k components are independent, one can compute the determinant of the resulting $k \times k$ correlation matrix (Anderson, 2003) and use permutation testing to establish significance. Guo and Modarres (2020) discuss testing the hypothesis of independence when $p > n$, propose a test statistic based on the correlation of the IPDS and compare it against several other novel statistics that exploit different properties of the sample interpoint distances. They show that the IPD correlation test detects linear, non-linear, monotone and non-monotone dependence structures.

3 Distribution of IPDs

In this section we study the distribution of the interpoint distances under the $\mathcal{M}$, normal, and elliptical models. While the asymptotic distribution of the interpoint distances is multivariate normal, we find their exact distribution for finite sample sizes when one can further assume that the form of the underlying distribution belongs to the following models.

3.1 The $\mathcal{M}$ Model

Srivastava and Kubokawa (2013) assume the following model, to which we refer as the $\mathcal{M}$ model. Suppose $\boldsymbol{\Sigma} = \mathbf{C}\mathbf{C}'$, where $\mathbf{C}$ is a $p \times p$ non-singular matrix. It is further assumed that $\mathbf{X}_i = \boldsymbol{\mu} + \mathbf{C}\mathbf{Z}_i$ where $\mathbb{E}(\mathbf{Z}_i) = \mathbf{0}$ and $\mathbb{C}\mathrm{ov}(\mathbf{Z}_i) = \mathbf{I}_p$ for $i = 1, \ldots, n$. For integers $\gamma_1, \ldots, \gamma_p$, $0 \leq \sum_{t=1}^p \gamma_t \leq 8$, $\mathbb{E}\Big[\prod_{t=1}^p z_{it}^{\gamma_t}\Big]) = \prod_{t=1}^p \mathbb{E}(z_{it}^{\gamma_t})$, where z_{it} is the tth component of the vector $\mathbf{z}_i = (z_{i1}, \ldots, z_{ip})'$. Hence, the model assumes existence of the moments of z_{it}, up to the eighth order. One can write the kurtosis $\gamma = \mathbb{E}\left(z_{it}^4\right) = K_4 + 3$ where K_4 is the excess coefficient of kurtosis defined by $K_4 = \gamma - 3$. The second model we consider is the *normal model*, for which $\gamma = 3$ so that $K_4 = 0$. In general, $K_4 \geq -2$

since $K_4 \geq \left(\frac{\mu_3}{\sigma^3}\right)^2 - 2$ where μ_3 is the third central moment. For any symmetric unimodal distribution, $K_4 \geq -\frac{6}{5}$. It is not difficult to show that K_4 is large for contaminated normal model and for heavy tailed distributions. Table 1 displays the values of K_4 for several univariate distributions.

Table 1: Excess Kurtosis K_4

U(0,1)	Beta(α,α)	N(0,1)	DE(0,1)	Logistic(0,1)	Exp(1)	t_v	LN(0,1)	$\Gamma(\alpha,\alpha)$
$-\frac{6}{5}$	$\frac{6}{2\alpha+3}$	0	$\frac{6}{5}$	3	6	$\frac{6}{v-4}, v > 4$	111	$3 + \frac{6}{\alpha}$

Lemma 3.1. *Suppose* $\mathbf{X} = \boldsymbol{\mu} + \mathbf{CZ}$ *is distributed under the* $\mathcal{M}$ *model. Let* $\boldsymbol{A}$ *be a* $p \times p$ *symmetric matrix. Let* $\boldsymbol{U} = (U_1, \ldots, U_p)'$ *and* $\boldsymbol{b} = (b_1, \ldots, b_p)'$ *where* $\mathbb{E}(\boldsymbol{U}) = \boldsymbol{0}$ *and* $\mathbb{C}\mathrm{ov}(\boldsymbol{U}) = \boldsymbol{I}_p$. *The distribution of the quadratic form is*

$$Q(\boldsymbol{X}) = \boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} = (\boldsymbol{U} + \boldsymbol{b})' \boldsymbol{\Lambda} (\boldsymbol{U} + \boldsymbol{b}) = \begin{cases} \sum_{t=1}^{p} \lambda_t (U_t + b_t)^2, & \text{if } \boldsymbol{\mu} \neq \boldsymbol{0}, \\ \sum_{t=1}^{p} \lambda_t U_t^2, & \text{if } \boldsymbol{\mu} = \boldsymbol{0}, \end{cases} \tag{3.1}$$

where $\boldsymbol{\Lambda} = diag(\lambda_1, \ldots, \lambda_p)$, *and* $\lambda_1, \ldots, \lambda_p$ *are the eigenvalues of* $\mathbf{CAC}$.

The random variables U_t includes standardized variables that appear in Table 1, with varying degrees of kurtosis. For example, when $\boldsymbol{\mu} = \mathbf{0}$, the distribution of IPDs is a linear combination of χ^2 random variables with one degree of freedom under the multivariate normal distribution. It is not difficult to show that $\mathbb{E}\left[Q(\mathbf{X})\right] = tr\,(\boldsymbol{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}$.

3.2 The Normal Model

It follows from (3.1) that, $d^2\left(\mathbf{X}_i, \mathbf{X}_j\right) \sim \sum_{t=1}^{p} \lambda_t \chi^2(1)$, where $\lambda_1, \ldots, \lambda_p$ are eigenvalues of $2\boldsymbol{\Sigma}$ and $\mathbf{U} = \mathbf{P}'\,(2\boldsymbol{\Sigma})^{-\frac{1}{2}}\left(\mathbf{X}_i - \mathbf{X}_j\right)$, $\mathbf{P}$ is an orthogonal matrix such that $\mathbf{P}'\,(2\boldsymbol{\Sigma})\,\mathbf{P} = diag(\lambda_1, \ldots, \lambda_p)$. Modarres and Song (2020) show that $\mathbb{E}\left(d^2(\mathbf{X}_i, \mathbf{X}_j)\right) = 2tr\,(\boldsymbol{\Sigma})$, and $\mathbb{V}\mathrm{ar}\left(d^2(\mathbf{X}_i, \mathbf{X}_j)\right) = 8tr\left(\boldsymbol{\Sigma}^2\right)$, respectively. A similar argument to the proof of Theorem 1 shows that the covariance of any two dependent IPDs is $2tr\left(\boldsymbol{\Sigma}^2\right)$. Gupta and Huang (2002) obtain the moment generating function of a quadratic form in Skew-normal random variables and prove that it is the same as that of the multivariate normal distribution. Hence, the distribution of the squared IPD remains true for

the skew-normal distributions. Lemma 3.1 also applies to any component of $\mathbf{X}_i^{(r)}$ of $\mathbf{X}_i = (\mathbf{X}_i^{(1)}, \mathbf{X}_i^{(2)}, \ldots, \mathbf{X}_i^{(k)})'$ for $i = 1, 2$ and $1 \leq r \leq k$. Hence, $\mathbf{X}_i^{(r)}$ is distributed as $N_{p_r}\left(\boldsymbol{\mu}^{(r)}, \boldsymbol{\Sigma}_{rr}\right)$ and $d^2\left(\mathbf{X}_i^{(r)}, \mathbf{X}_j^{(r)}\right) \sim \sum_{t=1}^{p_r} \lambda_t \chi^2(1)$ with mean $2tr\left(\boldsymbol{\Sigma}_{rr}\right)$ and variance $8tr\left(\boldsymbol{\Sigma}_{rr}^2\right)$ where $\lambda_{p_1}, \ldots, \lambda_{p_r}$ are eigenvalues of $2\boldsymbol{\Sigma}_{rr}$. The covariance between any two dependent IPDs in block r is $2tr\left(\boldsymbol{\Sigma}_{rr}^2\right)$, for $i, j, t \in \{1, \ldots, n\}$, and their correlation is $\frac{1}{4}$. From definition of IPD it follows that $d^2\left(\mathbf{X}_i, \mathbf{X}_j\right)$ is distributed as $\sum_{r=1}^{k} \sum_{t=1}^{p_r} \lambda_{tr} \chi^2(1)$ where λ_{tr} is the t-th eigenvalue of $\boldsymbol{\Sigma}_{rr}$.

3.3 The Elliptical Model

While assumptions of the $\mathcal{M}$ model are weak, it requires a linear dependence structure, which excludes several important distributions. In particular, the $\mathcal{M}$ model excludes many distributions from the elliptical family (Fang and Zhang, 1990) as distributions with non-linear correlations can not be modeled by the $\mathcal{M}$ model. The elliptical distributions (Keller, 1970) generalize multivariate normal distributions and contain a rich class of distributions that include both heavy and light tailed distributions such as symmetric stable distributions, the multivariate Cauchy, Student, logistic, generalized hyperbolic, Kotz, and symmetric Pearson type–VII distributions. A $p \times 1$ random vector $\mathbf{X}$ follows an elliptical distribution $E_p(\mu, \boldsymbol{\Sigma}, \psi)$ where ψ is a non-negative scalar function, if its characteristic function is of the form $\exp(i\mathbf{t}'\boldsymbol{\mu})\psi(\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t})$. In particular, if $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}_p$, then $\mathbf{X}$ has a spherically symmetric distribution, denoted by $\mathbf{X} \sim \mathbb{S}_p(\psi)$, where $\mathbb{S}_p$ is the surface of the unit sphere in $\mathbb{R}^p$. If $\mathbf{X} \sim E_p(\mu, \boldsymbol{\Sigma}, \psi)$ and $\boldsymbol{\Sigma} = \mathbf{BB}'$ is positive definite, then $\mathbf{T} = \boldsymbol{\Sigma}^{-1/2}\left(\mathbf{X} - \boldsymbol{\mu}\right) \sim \mathbb{S}_p(\psi)$. Furthermore, spherical distributions are invariant under orthogonal transformations. For any orthogonal matrix $\mathbf{O}$, the transformed vector $\mathbf{O}X$ and $\mathbf{X}$ are identically distributed as $\mathbb{S}_p(\psi)$.

Elliptical distributions are characterized by a kurtosis parameter κ that relates $\sigma_{ijkl} = \mathbb{E}[(\mathbf{X} - \mu)_i(\mathbf{X} - \mu)_j(\mathbf{X} - \mu)_k(\mathbf{X} - \mu)_l$ to $\sigma_{ij} = \mathbb{E}[(\mathbf{X} - \mu)_i(\mathbf{X} - \mu)_j]$ by $\sigma_{ijkl} = (\kappa + 1)(\sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk})$. Since $\sigma_{iiii} = 3(\kappa + 1)\sigma_{ii}^2$, elliptical distributions have the same marginal kurtoses $\kappa = \frac{1}{3}\left(\frac{\sigma_{iiii}}{\sigma_{ii}^2} - 3\right)$. The kurtosis κ does not reduce to the univariate kurtosis defined by K_4. All elliptical distributions are members of a location-scale family defined through an underlying spherical distribution. The elliptical distributions are closed under affine transformations and have conditional and marginal distributions that are also elliptically distributed. When $p = 1$ the class of elliptical distributions coincides with the class of univariate symmetric distributions. If $\boldsymbol{\Sigma}$ is a diagonal matrix, then the components of $\mathbf{X}$ are uncorrelated, but not independent unless the underlying

distribution is multivariate normal. The matrix $\mathbf{\Sigma}$ is proportional to the covariance matrix of $\mathbf{X}$, i.e., $\mathbb{C}\mathrm{ov}(\mathbf{X}) = \mathrm{a}\mathbf{\Sigma}$ where $a = -2\psi'(0)$.

Lemma 3.2. *Suppose* $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *be i.i.d. vectors drawn from an elliptical distribution* $E_p(\mu, \mathbf{\Sigma}, \psi)$. *The distribution of the IPD* $d^2(\boldsymbol{X}_1, \boldsymbol{X}_2)$ *is given by*

$$d^2(\boldsymbol{X}_1, \boldsymbol{X}_2) = \mathbf{U}' diag(\lambda_1, \ldots, \lambda_p)\mathbf{U} = \sum_{j=1}^{p} \lambda_j U_j^2, \tag{3.2}$$

where U_j *has a univariate symmetric distribution with center zero.*

4 Average Interpoint Distance

Let the sample average of all squared IPDs be denoted by

$$\bar{d}_p^{\,2} = \frac{1}{m} \sum_{1 \le i < j \le n} d^2\left(\mathbf{X}_i, \mathbf{X}_j\right), \tag{4.1}$$

and let the sample average of all squared IPDs for the r-th component, $r = 1, \ldots, k$, be denoted by

$$\bar{d}_{p_r}^{\,2} = \frac{1}{m} \sum_{1 \le i < j \le n} d^2\left(\mathbf{X}_i^{(r)}, \mathbf{X}_j^{(r)}\right). \tag{4.2}$$

Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i = \left(\bar{X}^{(1)}, \ldots, \bar{X}^{(k)}\right)$ and $\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X}_i - \bar{X})'(\mathbf{X}_i - \bar{X}) = (\mathbf{S}_{rt})$, for $r, t = 1, \ldots k$, denote the sample mean and covariance, respectively. In wide range of applications one is required to estimate the trace of a covariance matrix. To compute $\bar{d}_p^{\,2}$, one needs to find $m = \frac{n(n-1)}{2}$ IPDs so the computational complexity is $O\left(n^2\right)$. One can compute $\bar{d}_p^{\,2}$ with a lower computational complexity. It is straightforward to show that

$$\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{X}_i - \bar{X})(\mathbf{X}_i - \bar{X})' = \frac{1}{n(n-1)} \sum_{1 \le i < j \le d} \left(\mathbf{X}_i - \mathbf{X}_j\right)\left(\mathbf{X}_i - \mathbf{X}_j\right)'. \tag{4.3}$$

The identity follows by adding $\pm\bar{\boldsymbol{X}}$ to each term on the right hand side and

simplifying. It follows that

$$
\begin{aligned}
tr\,(\mathbf{S}) &= tr\left[\frac{1}{n(n-1)}\sum_{1\le i<j\le n}\left(\mathbf{X}_i-\mathbf{X}_j\right)\left(\mathbf{X}_i-\mathbf{X}_j\right)'\right] = \frac{1}{2}\bar{d}_p^{\,2},\\
tr\,(\mathbf{S}_{rr}) &= tr\left[\frac{1}{n(n-1)}\sum_{1\le i<j\le n}\left(\mathbf{X}_i^{(r)}-\mathbf{X}_j^{(r)}\right)\left(\mathbf{X}_i^{(r)}-\mathbf{X}_j^{(r)}\right)'\right] = \frac{1}{2}\bar{d}_{p_r}^{\,2}.
\end{aligned}
\tag{4.4}
$$

where $\mathbf{S}_{rr}$ denote the sample covariance matrix of the r-th block. It follows from (4.4) that $\bar{d}_p^{\,2} = 2tr(\mathbf{S})$ and $\bar{d}_{p_r}^{\,2} = 2tr(\mathbf{S}_{rr})$. The expected value of the average IPD is $\mathbb{E}\left(\bar{d}_p^{\,2}\right) = 2tr\,(\mathbf{\Sigma})$. The computational complexity to find $\bar{d}_p^{\,2}$ is $O(np)$. This computational complexity is smaller than $O\left(n^2\right)$ unless p is larger than n.

Consider an affine transformation $\mathbf{Y} = \mathbf{BX}+\mathbf{b}$ where $\mathbf{B}$ is a nonsingular $p\times p$ matrix of constants and $\mathbf{b}$ is a $p\times 1$ vector. Since $\mathbf{S}_y = \mathbf{BS}_x\mathbf{B}$, one can verify that $tr(\mathbf{S}_y) = tr(\mathbf{AS})$ and $tr(\mathbf{S}_y^2) = tr(\mathbf{ASAS})$ where $A = \mathbf{B'B}$. When $\mathbf{B}$ is orthogonal, $\mathbf{A} = \mathbf{I}_p$. Hence, orthogonal transformations do not alter IPDs.

Examples 4.1. Consider i.i.d. vectors $\{\mathbf{X}_i\}_{i=1}^n$ drawn from a p-variate lognormal distribution with parameters $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$. It is well-known that $\mathbb{V}\mathrm{ar}(x_t) = \exp(2\boldsymbol{\mu}_t + \sigma_{tt})(\exp(\sigma_{tt})-1)$ where x_t is the t-th component of $\mathbf{X}_i$. One can show that $\bar{d}_p^{\,2} = 2tr\,(\mathbf{S}) = 2\sum_{t=1}^p \exp(2\bar{x}_t + s_{tt})(\exp(s_{tt})-1)$ where $\bar{\mathbf{x}}$ and $\mathbf{S} = (s_{ij})$ are parameter estimates of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$, respectively. When $\mathbf{X} \sim E_p(\mu, \mathbf{\Sigma}, \psi)$, one has $\mathbb{C}\mathrm{ov}(\mathbf{X}) = \mathrm{a}\mathbf{\Sigma}$, so that $\bar{d}_p^{\,2} = 2a\ tr\,(\mathbf{S})$ where $a = -2\psi'(0)$. Hence, given the form of the covariance matrix of a distribution, (4.4) allows one to compute the average squared IPD without actually computing the IPDs.

Let $\bar{d}_{(t)}^{\,2}$ denote the IPDs of the t-th univariate component of $\mathbf{X}$ and r_{ts} be the sample correlation between the t-th and s-th univariate components of $\mathbf{X}$. Since $2s_{tt} = \bar{d}_{(t)}^{\,2}$ one obtains $tr(\mathbf{S}^2) = \frac{1}{4}\sum_{t=1}^p \bar{d}_{(t)}^{\,4} + \frac{1}{2}\sum_{1\le t<s\le p} r_{ts}^2\bar{d}_{(t)}^{\,2}\bar{d}_{(s)}^{\,2}$.

Lemma 4.1. *Let $\le_{St}$ be the stochastic ordering between two random variables. Using (4.4), the following inequalities hold:*

- $\bar{d}_p^{\,2} \le_{St} 2\sqrt{p\ tr(\mathbf{S}^2)}$ *and*
- $\bar{d}_{p_r}^{\,2} \le_{St} 2\sqrt{p_r\ tr(\mathbf{S}_{rr}^2)}$ *for* $r = 1, \ldots, k$.

5 Asymptotic Distribution of IPDs

It follows from Theorem 1 of Li (2018) and the stated regularity conditions that as p approaches infinity and for fixed n, $\frac{d^2_{(X)}-2tr(\mathbf{\Sigma})}{\sqrt{\mathbb{V}\mathrm{ar}\left(\mathrm{d}^2_{(X)}\right)}}$ converges is distribution to $N(0,1)$ where $\mathbb{V}\mathrm{ar}\left(\mathrm{d}^2_{(X)}\right)$ depends on F. Let $\mathbf{R} = (d^2(\mathbf{X}_1,\mathbf{X}_2), d^2(\mathbf{X}_1,\mathbf{X}_3),\ldots,d^2(\mathbf{X}_{n-1},\mathbf{X}_n))'$ denote the m-dimensional vector of all IPDs. It follows that $\mathbb{E}(\mathbf{R}) = 2tr\,(\mathbf{\Sigma})\,\mathbf{1}'_m$ where $\mathbf{1}_m$ is a m-dimensional vector of 1s. Under the *normal model* the covariance $\mathbf{\Sigma}_R$ is a $m\times m$ matrix with main diagonal $8tr\left(\mathbf{\Sigma}^2\right)$. The off-diagonal elements of $\mathbf{\Sigma}_R$ are either $2tr\left(\mathbf{\Sigma}^2\right)$ if the two IPDs have an index in common or zero, otherwise. Hence, ${\mathbf{\Sigma}_R}^{-1/2}\,(\mathbf{R}-\mathbb{E}(\mathbf{R}))$ converges in distribution to the standard multivariate normal normal distribution as $p\to\infty$ for fixed n. Similarly, as $p\to\infty$ such that p_r approaches infinity, $d^2_{(X^{(r)})}$ converges in distribution to a normal distribution with mean $2tr\,(\mathbf{\Sigma}_{rr})$ and variance $8tr\left(\mathbf{\Sigma}^2_{rr}\right)$ for $r=1,\ldots,k$.

5.1 The Elliptical Model

Muirhead (1982) proves that under the *elliptical model* $\sqrt{n-1}\,(\mathbf{S}-\mathbf{\Sigma})$ converges to a $p(p+1)/2$ dimensional normal distribution with zero mean and covariance $\mathbf{\Gamma} = (\gamma_{ij,kl})$ as $n\to\infty$ where $\gamma_{ij,kl} = \mathbb{C}ov(s_{ij},s_{kl}) = (\kappa+1)(\sigma_{ij}\sigma_{kl}+\sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk})$. It follows that $\sqrt{n-1}\,(diag(\mathbf{S})-diag(\mathbf{\Sigma}))$ converges to a p dimensional normal distribution with zero means, variances $\gamma_{ii,ii} = \mathbb{V}ar(s_{ii}) = 3(\kappa+1)\sigma^2_{ii}$ and covariances $\gamma_{ii,kk} = \mathbb{C}ov(s_{ii},s_{kk}) = (\kappa+1)(\sigma_{ii}\sigma_{kk}+2\sigma^2_{ik})$ as $n\to\infty$.

Following Iwashita and Siotani (1994), let $h_j(\mathbf{S})$ for $j=1,2$ be continuous and scalar-valued functions of $\mathbf{S}$ with continuous first and second order derivatives in a neighborhood of $\mathbf{S}=\mathbf{\Sigma}$. It follows that $\sqrt{n}(h_j(\mathbf{S})-h_j(\mathbf{\Sigma}))$ are asymptotically normally distributed with mean 0 as $n\to\infty$. Using the delta method and (4.4) one can show that as $n\to\infty$,

$$\sqrt{n-1}\left(\bar{d}^2_p-2tr(\mathbf{\Sigma})\right)\to N\left(0,8(1+\kappa)tr\left(\mathbf{\Sigma}^2\right)+4\kappa\,(tr\mathbf{\Sigma})^2\right). \tag{5.1}$$

Similarly, one can show that for $r\neq t=1,\ldots,k$,

$$\mathbb{C}\mathrm{ov}(\bar{d}^2_{p_r},\bar{d}^2_{p_t}) = \frac{8(1+\kappa)}{n-1}tr\left(\mathbf{\Sigma}'_{rt}\mathbf{\Sigma}_{rt}\right)+\frac{4\kappa}{n-1}tr(\mathbf{\Sigma}_{rr})\;tr(\mathbf{\Sigma}_{tt}), \tag{5.2}$$

It follows that as $n\to\infty$,

$$\sqrt{n-1}\left(\bar{d}^2_{p_r}-2tr(\mathbf{\Sigma}_{rr})\right)\to N\left(0,8(1+\kappa)tr\left(\mathbf{\Sigma}^2_{rr}\right)+4\kappa(tr\mathbf{\Sigma}_{rr}\right)^2). \tag{5.3}$$

Notice that $\mathbf{\Sigma}_{rt} = \mathbf{0}$ implies $\mathbb{C}\mathrm{ov}(\bar{\mathrm{d}}_{\mathrm{p_r}}^2, \bar{\mathrm{d}}_{\mathrm{p_t}}^2) = \frac{4\kappa}{\mathrm{n-1}}\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{rr}})\,\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{tt}}) > 0$ unless $\kappa = 0$. Hence, Theorem 2.1 cannot be extended to the elliptical model. It follows that $\mathbb{E}(\bar{d}_p^2) = \eta = 2tr(\mathbf{\Sigma})$ and $\mathbb{E}(\bar{d}_{p_r}^2) = \eta_r = 2tr(\mathbf{\Sigma}_{rr})$. Under the *normal model*, when $\kappa = 0$, $\mathbb{V}\mathrm{ar}(\bar{d}_p^2)) = \mathbb{V}\mathrm{ar}(\sum_{r=1}^k 2tr(S_{rr})) = \frac{8}{n-1}tr\left(\mathbf{\Sigma}^2\right)$ and $\mathbb{V}\mathrm{ar}(\bar{d}_{p_r}^2)) = \frac{8}{n-1}tr\left(\mathbf{\Sigma}_{rr}^2\right)$.

5.2 The $\mathcal{M}$ Model

Under this model, $\mathbf{X}_i = \boldsymbol{\mu} + \mathbf{C}\mathbf{Z}_i$, for $i = 1, \ldots, n$ where $\mathbf{\Sigma} = \mathbf{C}\mathbf{C}'$ and $\mathbf{C}$ is a $p \times p$ non-singular matrix. It follows that $\mathbb{E}(\mathbf{Z}_i) = \mathbf{0}$, $\mathbb{C}\mathrm{ov}(\mathbf{Z}_i) = \mathbf{I}_p$, and $\mathbb{E}\left(z_{it}^4\right) = K_4 + 3$ where K_4 is the excess coefficient of kurtosis and z_{it} is the tth component of the vector $\mathbf{z}_i = (z_{i1}, \ldots, z_{ip})'$. Hence, $d^2(\mathbf{X}_i, \mathbf{X}_j) = (\mathbf{Z}_i - \mathbf{Z}_j)'\mathbf{\Sigma}(\mathbf{Z}_i - \mathbf{Z}_j)$. Using Lemma 6.1 of Srivastava and Kubokawa (2013), one can show that

$$\mathbb{V}\mathrm{ar}\left(d^2(\mathbf{X}_i, \mathbf{X}_j)\right) = 4K_4 \sum_{r=1}^{p} \sigma_{rr}^2 + 8tr\left(\mathbf{\Sigma}^2\right), \tag{5.4}$$

which reduces to $8tr(\mathbf{\Sigma}^2)$ when $K_4 = 0$ under the *normal model*. Note that when $d^2(\mathbf{X}_i, \mathbf{X}_j)$ and $d^2(\mathbf{X}_i, \mathbf{X}_k)$ have no index in common their covariance is zero. For $i \neq j \neq k = 1, \ldots, n$, we obtain

$$\mathbb{C}\mathrm{ov}\left(d^2(\mathbf{X}_i, \mathbf{X}_j), d^2(\mathbf{X}_i, \mathbf{X}_k)\right) = K_4 \sum_{r=1}^{p} \sigma_{rr}^2 + 2tr(\mathbf{\Sigma}^2). \tag{5.5}$$

Note that $\mathbb{C}\mathrm{orr}\left(d^2(\mathbf{X}_i, \mathbf{X}_j), d^2(\mathbf{X}_i, \mathbf{X}_k)\right) = 1/4$, which is the same as the correlation under normality. To find the variance of $\mathbb{V}\mathrm{ar}(\bar{\mathrm{d}}_{\mathrm{p}}^2)$, using (4.1), we have

$$m^2\mathbb{V}\mathrm{ar}(\bar{d}_p^2) = m\mathbb{V}\mathrm{ar}\left(d^2(\mathbf{X}_i, \mathbf{X}_j)\right) + 2m(n-2)\,\mathbb{C}\mathrm{ov}\left(d^2(\mathbf{X}_i, \mathbf{X}_j), d^2(\mathbf{X}_i, \mathbf{X}_k)\right). \tag{5.6}$$

Hence, the variance of the average squared IPD is

$$\mathbb{V}\mathrm{ar}(\bar{d}_p^2) = \frac{4}{n-1}\left(K_4 \sum_{i=1}^{p} \sigma_{ii}^2 + 2tr\left(\mathbf{\Sigma}^2\right)\right). \tag{5.7}$$

Similarly, one can show that for $r \neq t = 1, \ldots, k$,

$$\mathbb{C}\mathrm{ov}(\bar{d}_{p_r}^2, \bar{d}_{p_t}^2) = \frac{4}{n-1}\left(K_4 \sum_{i=1}^{p} \sigma_{(r)ii}^2 + 2tr(\mathbf{\Sigma}_{tr}'\mathbf{\Sigma}_{rt})\right), \tag{5.8}$$

$$\mathbb{V}\mathrm{ar}(\bar{d}_{p_r}^{\,2}) = \frac{4}{n-1}\left(K_4 \sum_{i=1}^{p_r} \sigma^2_{(r)ii} + 2tr\left(\mathbf{\Sigma}^2_{rr}\right)\right). \tag{5.9}$$

where $\sigma^2_{(r)ii}$ is the i-th diagonal element of $\mathbf{\Sigma}_{rr}$, for $r = 1, \ldots, k$. Notice that $\mathbf{\Sigma}_{rt} = \mathbf{0}$ implies $\mathbb{C}\mathrm{ov}(\bar{\mathrm{d}}^{\,2}_{\mathrm{p_r}}, \bar{\mathrm{d}}^{\,2}_{\mathrm{p_t}}) = \frac{4\mathrm{K}_4}{\mathrm{n}-1}\sum_{\mathrm{i}=1}^{\mathrm{p_r}} \sigma^2_{(\mathrm{r})\mathrm{ii}} > 0$ unless $K_4 = 0$. Hence, Theorem 2.1 cannot be extended to the $\mathcal{M}$ model. Under the $\mathcal{M}$ model, the average squared IPD converges in distribution to a normal distribution. As $n \to \infty$,

$$\sqrt{n-1}(\bar{d}_p^{\,2} - \eta) \to N(0, 4K_4 \sum_{i=1}^{p} \sigma^2_{ii} + 8tr\left(\mathbf{\Sigma}^2\right)), \tag{5.10}$$

$$\sqrt{n-1}(\bar{d}_{p_r}^{\,2} - \eta_r) \to N(0, 4K_4 \sum_{i=1}^{p_r} \sigma^2_{(r)ii} + 8tr\left(\mathbf{\Sigma}^2_{rr}\right)). \tag{5.11}$$

To examine the rate of convergence of the average squared interpoint distance to normality as a function of n and p we replicated the following experiment 1000 times. We generated samples of size $n = \{20, 50, 100, 200, 300\}$ and $p = \{2, 10, 50, 100, 200, 300, 500, 1000\}$ from multivariate normal and absolute value of a multivariate normal distribution with constant correlation structure $\mathbf{\Sigma} = (\rho_{ij}) = (0.0)(0.5)$. We applied the Shapiro-Wilk test of normality to the average squared IPD. Tables 2-3 show the type I error of testing whether the average squared IPDs follow a normal distribution at nominal level $\alpha = 0.05$. Table entries at (n, p) that are close to 0.05 indicate convergence to normality has taken hold. Type I-error rates converge to the nominal level faster under independence.

Table 2: Type-I error for testing the asymptotic normality of the average squared IPD when $\mathbf{X} \sim \mathbf{N}(\mathbf{0}, \mathbf{\Sigma})$ and $\mathbf{\Sigma}$ has constant correlation $\rho = 0.0$ (first cell entry) and $\rho = 0.5$ (second cell entry).

$n \mid p$	2	10	50	100	200	300	500	1000
20	.12 .17	.10 .16	.10 .16	.04 .18	.04 .18	.04 .26	.06 .24	.06 .15
50	.08 .07	.05 .05	.02 .07	.04 .09	.03 .11	.05 .10	.06 .12	.07 .07
100	.09 .06	.05 .12	.03 .10	.01 .09	.08 .03	.05 .08	.04 .08	.03 .06
200	.01 .02	.02 .11	.03 .09	.04 .10	.02 .09	.06 .05	.01 .06	.05 .08
300	.04 .06	.04 .09	.03 .07	.05 .04	.05 .03	.02 .05	.01 .07	.04 .05

Table 3: Type-I error for testing the asymptotic normality of the average squared IPD when $\mathbf{X} \sim |\mathbf{N}(\mathbf{0}, \mathbf{\Sigma})|$ and $\mathbf{\Sigma}$ has constant correlation $\rho = 0.0$ (first cell entry) and $\rho = 0.5$ (second cell entry).

$n \backslash p$	2	10	50	100	200	300	500	1000
20	.20 .25	.12 .35	.05 .26	.04 .31	.06 .19	.06 .28	.03 .26	.06 .26
50	.17 .28	.08 .13	.04 .12	.06 .08	.02 .07	.09 .08	.09 .08	.07 .07
100	.14 .11	.06 .07	.03 .07	.02 .10	.03 .10	.04 .07	.07 .09	.07 .06
200	.08 .06	.06 .06	.05 .08	.08 .08	.05 .07	.07 .09	.09 .05	.03 .06
300	.08 .08	.07 .05	.04 .05	.05 .10	.05 .05	.03 .04	.06 .04	.07 .06

6 Vector Correlation

Consider vectors $\mathbf{X}^{(r)}$ with p_r variables and $\mathbf{X}^{(t)}$ with p_t variables and the corresponding squared average IPD $\bar{d}^2_{p_r}$ and $\bar{d}^2_{p_t}$ for $r \neq t = 1, \ldots, k$. Consider the $\mathcal{M}$ model and to emphasize the dependence on K_4 let $\psi_{(K_4)(rt)} = \mathbb{C}\mathrm{orr}(\bar{d}^2_{p_r}, \bar{d}^2_{p_t})$ where $0 \leq \psi_{(K_4)(rt)} \leq 1$. Using (5.7) and (5.8) one can show that

$$\psi_{(K_4)(rt)} = \frac{K_4 \sum_{i=1}^{p} \sigma^2_{(r)ii} + 2tr\left(\mathbf{\Sigma}'_{tr}\mathbf{\Sigma}_{rt}\right)}{\left(K_4 \sum_{i=1}^{p_r} \sigma^2_{(r)ii} + 2tr\left(\mathbf{\Sigma}^2_{rr}\right)\right)^{1/2}\left(K_4 \sum_{i=1}^{p_t} \sigma^2_{(t)ii} + 2tr\left(\mathbf{\Sigma}^2_{tt}\right)\right)^{1/2}}. \tag{6.1}$$

When $K_4 = 0$, under a *normal model*, the population correlation coefficient between $\bar{d}^2_{p_r}$ and $\bar{d}^2_{p_t}$ in (6.1) reduces to

$$\psi_{(0)(rt)} = \mathbb{C}\mathrm{orr}(\bar{\mathrm{d}}^2_{\mathrm{p_r}}, \bar{\mathrm{d}}^2_{\mathrm{p_t}}) = \frac{\mathrm{tr}(\mathbf{\Sigma}'_{\mathrm{rt}}\mathbf{\Sigma}_{\mathrm{tr}})}{\sqrt{tr(\mathbf{\Sigma}^2_{rr})tr(\mathbf{\Sigma}^2_{tt})}}. \tag{6.2}$$

It is clear that $\psi_{(0)(rt)} = 0$ if and only if $\mathbf{\Sigma}_{rt} = \mathbf{0}$. Similarly, consider the *elliptical model* and to emphasize the dependence on κ let $\psi^*_{(\kappa)(rt)} = \mathbb{C}\mathrm{orr}(\bar{d}^2_{p_r}, \bar{d}^2_{p_t})$ where $0 \leq \psi^*_{(\kappa)(rt)} \leq 1$. Using (5.2) and (5.3) one obtains

$$\psi^*_{(\kappa)(rt)} = \frac{8(1+\kappa)tr\left(\mathbf{\Sigma}'_{rt}\mathbf{\Sigma}_{rt}\right) + 4\kappa\ tr(\mathbf{\Sigma}_{rr})tr(\mathbf{\Sigma}_{tt})}{[8(1+\kappa)tr\left(\mathbf{\Sigma}^2_{rr}\right) + 4\kappa\ (tr\mathbf{\Sigma}_{rr})^2]^{1/2}[8(1+\kappa)tr\left(\mathbf{\Sigma}^2_{tt}\right) + 4\kappa\ (tr\mathbf{\Sigma}_{tt})^2]^{1/2}}. \tag{6.3}$$

When $\kappa = 0$, under a *normal model*, the population correlation coefficient between $\bar{d}^2_{p_r}$ and $\bar{d}^2_{p_t}$ in (6.3) reduces (6.2) so that $\psi^*_{(0)(rt)} = \psi_{(0)(rt)}$. It is clear that $\psi^*_{(0)(rt)} \neq 0$ unless $\kappa = 0$.

In fact, (6.2) is the vector correlation coefficient $RV(\mathbf{X}^{(r)}, \mathbf{X}^{(t)})$ defined by Escoufier (1973) and Robert and Ranger (1985). Hence, under multivariate normal distribution, the RV coefficient between two random vectors is the same as the correlation of their interpoint distances. Escoufier (1973) shows that if $p_r = p_t = 1$, then $RV = r^2$, the square of the simple correlation coefficient. Furthermore, $0 \leq RV \leq 1$ and $RV = 0$ if and only if $\mathbf{\Sigma}_{rt} = \mathbf{0}$.

Examples 6.1. Let $\lambda_1, \ldots, \lambda_p$ are eigenvalues of $\mathbf{\Sigma}$. It follows from the Cauchy-Schwarz inequality that $\left(\sum_{i=1}^p \lambda_i^t\right)^2 \leq p\left(\sum_{i=1}^p \lambda_i^{2t}\right)$. Since $tr\left(\mathbf{\Sigma}^t\right) = \sum_{i=1}^p \lambda_i^t$ for $t \geq 1$ we obtain $\left(tr\mathbf{\Sigma}\right)^2 \leq p\ tr\left(\mathbf{\Sigma}^2\right)$, so that $\theta = \frac{p(tr\mathbf{\Sigma}^2)}{(tr\mathbf{\Sigma})^2} \geq 1$ with equality holding if and only if $\lambda_1 = \lambda_2 = \ldots, \lambda_p = \lambda$, for some constant λ. Srivastava (2005) and others have used an estimate of θ to construct tests of sphericity; i. e. $\mathbf{\Sigma} = \lambda I_p$. The squared coefficient of variation for the IPD random variable $d^2_{(X)}$ is defined as $c_v^2 = \frac{\mathbb{V}ar(d^2_{(X)})}{(\mathbb{E}(d^2_{(X)}))^2}$. It is not difficult to show that under a *normal model* $c_v^2 = 2\theta/p$. Hence, the squared coefficient of variation of IPDs is a linear function of the measure for the sphericity of the underlying *normal model*.

The distance concentration phenomenon (Francois et al., 2007) describes the effect of high dimension p on IPDS. Under appropriate moment assumptions, one can show that the coefficient of variation of the IPDs tends to zero as p tends to infinity. That is $c_v = \sqrt{2\theta}/\sqrt{p}$ tends to zero as $p \to \infty$. As a result of the distance concentration phenomenon one can show that IPDs of vectors with dependent components are less concentrated (more variable) than IPDs of vectors with independent components.

Consider the null hypothesis $G_0 : \mathbb{C}\mathrm{ov}(d^2_{(X^{(r)})}, d^2_{(X^{t)})}) = 0$. Let $G_a : \mathbb{C}\mathrm{ov}(d^2_{(X^{(r)})}, d^2_{(X^{t)})}) \neq 0$ for at least one $r \neq t = 1, \ldots, k$ be the alternative hypothesis. Under the $\mathcal{M}$ model, we define the ratio of the variance under G_a to the variance under G_0 by

$$\beta_{K_4} = \frac{\mathbb{V}\mathrm{ar}_{\mathrm{G_a}}(\bar{\mathrm{d}}^2_{\mathrm{p}})}{\mathbb{V}\mathrm{ar}_{\mathrm{G_0}}(\bar{\mathrm{d}}^2_{\mathrm{p}})} = \frac{K_4 \sum_{i=1}^p \sigma_{ii}^2 + 2tr\left(\mathbf{\Sigma}^2\right)}{K_4 \sum_{i=1}^p \sigma_{ii}^2 + 2\sum_{r=1}^k tr\left(\mathbf{\Sigma}_{rr}^2\right)}. \tag{6.4}$$

Using (5.1), one obtains a similar expression for the variance ratio under the *elliptical model*. Under the *normal model*, the variance ratio reduces to $\beta_0 = \frac{tr(\mathbf{\Sigma}^2)}{\sum_{r=1}^k tr(\mathbf{\Sigma}_{rr}^2)}$. Srivastava

(2005) constructed a statistic for testing the hypothesis that $\mathbf{\Sigma}$ is a diagonal covariance matrix when $n < p$ and investigates a distance function for testing block independence using $m_d = \beta_0 - 1$. It is clear that $\beta_0 \geq 1$ with equality holding when $\mathbf{\Sigma}$ is diagonal. He further shows that his proposed test statistic is normally distributed for large (n, p) and is robust when $K_4 = o(p^{-\epsilon})$ where $\epsilon > 0$.

Examples 6.2. Consider i.i.d. random vectors $\mathbf{X}_i \sim N_p(\mathbf{0}, \boldsymbol{\rho})$ for $i = 1, 2$ where $\boldsymbol{\rho}$ has 1 on the main diagonal, constant correlation ρ on the off-diagonal and $-\frac{1}{p-1} < \rho < 1$. It follows that under $G_0 : \boldsymbol{\rho} = \mathbf{I}$, we have $\mathbb{E}(d^2_{(X)}) = 2p$ and $\mathbb{V}\mathrm{ar}(\mathrm{d}^2_{(X)}) = 8p$ whereas under $G_a : \boldsymbol{\rho} \neq \mathbf{I}$, $\mathbb{E}(d^2_{(X)}) = 2p$ and $\mathbb{V}\mathrm{ar}(\mathrm{d}^2_{(X)}) = 8\sum_{i=1}^{p} \lambda_i^2 = 8p(1+(p-1)\rho^2)$ since $\lambda_1 = 1+(p-1)\rho$ and $\lambda_2 = \ldots, \lambda_p = 1 - \rho$ and the variance ratio is $\beta_0 = 1 + (p-1)\rho^2$. Note that $\beta_0 = 1$ when $\rho = 0$ (the covariance matrix is diagonal) and as ρ approaches 1, β_0 approaches p.

7 Comparing Total Variances

Let $\eta_r = \mathbb{E}(d^2_{(X^{(r)})})$ for $r = 1, \ldots, k$. We are also interested in testing the hypothesis L_0 in (1.1) that the block total variances are the same across the k dependent components against general alternative $L_a : \eta_r \neq \eta_t$ for at least one pair (r, t) where $r \neq t = 1, \ldots k$. Total variance (Anderson, 2003) is defined as the trace of a covariance matrix and a measure of global dispersion of a multivariate dataset . Denote the $k \times 1$ vector of the squared average distances for the r-th $(1 \leq r \leq k)$ components of $\mathbf{X}_i$ and $\mathbf{X}_j$ by $\mathbf{W} = \left(\bar{d}^2_{p_1}, \bar{d}^2_{p_2}, \ldots, \bar{d}^2_{p_k}\right)$. It follows that $\mathbf{\Psi}^{-1/2}\left(\mathbf{W} - \boldsymbol{\eta}_W\right)$ converges in distribution to a standard multivariate normal distribution as $p \to \infty$ where $\boldsymbol{\eta}_W = \mathbb{E}(\mathbf{W}) = (\eta_1, \eta_2, \ldots, \eta_k) = (2tr(\mathbf{\Sigma}_{11}), 2tr(\mathbf{\Sigma}_{22}), \ldots, 2tr(\mathbf{\Sigma}_{kk}))$. The covariance of W under the $\mathcal{M}$ model are given by (5.7) and (5.9) and under an *elliptical model* by (5.1) and (5.3). Under the *normal model* when $K_4 = 0$ or $\kappa = 0$ and we further obtain

$$\mathbf{\Psi} = \mathbb{C}\mathrm{ov}(\mathbf{W}) = \frac{8}{n-1}\begin{bmatrix} tr(\mathbf{\Sigma}_{11})^2 & tr\left(\mathbf{\Sigma}'_{12}\mathbf{\Sigma}_{12}\right) & \ldots tr\left(\mathbf{\Sigma}'_{1k}\mathbf{\Sigma}_{1k}\right) \\ tr\left(\mathbf{\Sigma}'_{21}\mathbf{\Sigma}_{21}\right) & tr(\mathbf{\Sigma}_{22})^2 & \ldots tr\left(\mathbf{\Sigma}'_{2k}\mathbf{\Sigma}_{k2}\right) \\ \vdots & \vdots & \vdots \\ tr\left(\mathbf{\Sigma}'_{k1}\mathbf{\Sigma}_{1k}\right) & tr\left(\mathbf{\Sigma}'_{k2}\mathbf{\Sigma}_{2k}\right) & \ldots tr(\mathbf{\Sigma}_{kk})^2 \end{bmatrix}. \tag{7.1}$$

Construct a $(k-1) \times k$ contrast matrix $\mathbf{A} = (a)_{rt}$ where $a_{rr} = 1$ and $a_{r(r+1)} = -1$ for $r = 1, \ldots k-1$ so that row sums are 0. It follows that $\mathbf{Y} = \mathbf{AW}$ has a $k-1$ dimensional multivariate normal distribution with mean $\boldsymbol{\mu}_y = \mathbf{A}\boldsymbol{\eta}_W$ and covariance $\mathbf{A\Psi A'}$. Hence, $T = \mathbf{Y}'(\mathbf{A}\hat{\mathbf{\Psi}}\mathbf{A}')^{-1}\mathbf{Y}$ has a χ^2 distribution with $k-1$ degrees of freedom as $p \to \infty$ where

$\hat{\mathbf{\Psi}}$ is a consistent estimate of $\mathbf{\Psi}$. Assuming a Wishart distribution, Srivastava (2005) in lemma 2.1 discusses consistent estimator of $tr(\mathbf{\Sigma}^i)$ for $i = 1, 2$, which appear in $\mathbf{\Psi}$.

7.1 Visualizing Block Variances

In order to visualize the total variances of the blocks and provide a basis for their comparison we propose the following method. Let $R = \max_{r=1}^{k}(\bar{d}_{p_r}^2) - \min_{r=1}^{k}(\bar{d}_{p_r}^2)$ denote the range of the average IPDs of all blocks. We will divide the range into s evaluation points denoted by $\delta(t)$ for $t = 1, \ldots s$. Denote the cumulative distributions of $\bar{d}_{p_r}^2$ evaluated at $\delta(t)$ by $H_r(t) = \mathbb{P}(\bar{d}_{p_r}^2 \leq \delta(t))$. Let $I(.)$ denote the indicator function and estimate the CDF by

$$\hat{H}_r(t) = \frac{1}{m} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} I(\bar{d}_{p_r}^2 \leq \delta(t)). \tag{7.2}$$

We obtain a simultaneous plot of $(\delta(t), \hat{H}_r(\delta(t)))$, for $t = 1, \ldots, s$, $r = 1, \ldots, k$.

Examples 7.1. To illustrate the simultaneous display, we use $s = 100$, $p = 500$ and $k = 4$, so that each block contains 125 variables. We generate $n = 30$ observations from $\mathbb{N}(\mathbf{0}, \mathbf{\Sigma})$ where $\mathbf{\Sigma}$ has constant correlation ρ. Figure 8 shows the simultaneous plots of the total variances when L_0 is true, $\sigma_{tt} = 1$ and $\rho = 0.5$ for $t = 1, \ldots, p$. As one expects, the ECDFs of the total variances concentrate in a narrow band under L_0. Figure 8 displays the simultaneous plot of (7.2) under multivariate normal distributions where variance of the block r is $r = 1, \ldots, 4$ and $\rho = 0.5$. Here, L_0 is false and we expect the ECDFs to differ. Moreover, the display provides a ranking of the block total variances.

8 Summary and Conclusion

We discuss blocks of $k \geq 2$ IPDs and show that under the normal and the power-normal family of distributions independence of blocks is equivalent to zero covariance between the blocks. We discuss the exact and asymptotic distribution of the IPDs and their average under the *normal*, *SM* and *elliptical models*. We show that correlation between two blocks of IPDs is equivalent to vector correlation. We obtain a relationship between the coefficient of variation of IPDs and a measure of sphericity. A sampling experiment determines when the average IPD converge to normality as a function of n and p. Finally, a test of the homogeneity of the total block variances and a simultaneous plot of the total variances are presented. Future research directions includes the study

the small sample behavior of the proposed tests, comparison of the power and type-one error rate against competing tests.

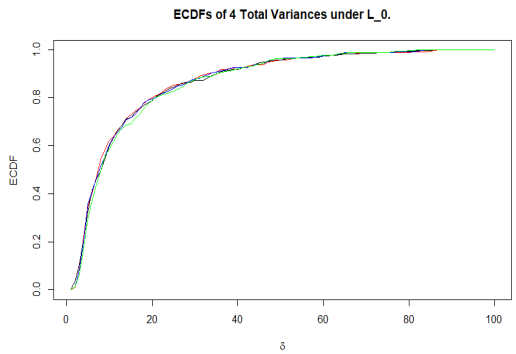

Figure 1: Simultaneous plots of the ECDF of 4 total variances where $\sigma_{rr} = 1$ for $r = 1, \ldots, 4$.

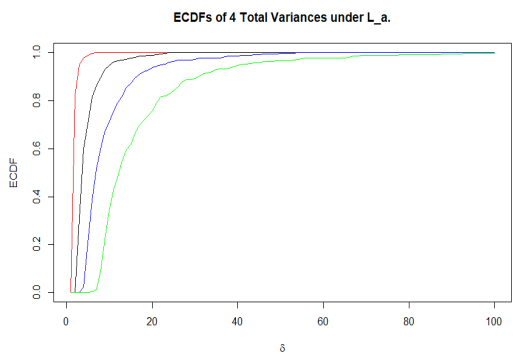

Figure 2: Simultaneous plots of the ECDF of 4 total variances where variance of the block r is $r = 1, \ldots, 4$.

References

Anderson, T. W. (2003), *An Introduction to Multivariate Statistical Analysis*. New Jersey: Wiley-Interscience.

Bottesch, T., Bühler, T. Kächele, M. (2016), Speeding up k-means by approximating Euclidean distances via block vectors. *Proceedings of the 33rd International Conference on International Conference on Machine Learning*, Volume **48**, 2578–2586.

Escoufier, Y. (1973), Le traitement des variables vectorielles. *Biometrics*, **29**, 751-760.

Fang, K. T., Zhang, Y. T. (1990), *Generalized multivariate analysis. Springer-Verlag*, Berlin; Science Press, Beijing.

Flexer, A., Schnitzer, D. (2015), Choosing l_p norms in high-dimensional spaces based on hub analysis. *Neurocomputing*, **169**, 281-287.

Freeman, J. and Modarres, R. (2005), Efficiency of test for independence after Box-Cox transformation. *Journal of Multivariate Analysis*, **95**, 107-118.

Freeman, J. and Modarres, R. (2006), Inverse Box-Cox: The power-normal distribution. *Statistics and Probability Letters*, **76**, 764-772.

Guo, L., Modarres, R. (2019), Interpoint Distance Classification of High Dimensional Discrete Observations. *International Statistical Review*, **87**(2), 191-206.

Guo, L., Modarres, R. (2020), Nonparametric tests of independence based on interpoint distance. *Journal of Nonparametric Statistics*, **32**(1), 225-245.

Gupta, A. K. and Huang, W. J. (2002), Quadratic forms in skew normal variates. *J. Math. Anal. Appl.*, **273**, 558-564.

Iwashita, T. and Siotani, M. (1994), Asymptotic Distributions of Functions of a Sample Covariance Matrix under the Elliptical Distribution. *The Canadian Journal of Statistics*, **22**(2), 273-283.

Li, J. (2018), Asymptotic normality of interpoint distances for high-dimensional data with applications to the two-sample problem. *Biometrika*, **105**(3), 529-546.

Marozzi, M. (2015), Multivariate multidistance tests for high-dimensional low sample size case-control studies. *Statistics in Medicine*, **34**, 1511–1526.

Marozzi, M. (2016), Multivariate tests based on interpoint distances with application to magnetic resonance imaging. *Stat. Methods Med. Res.*, **25**(6), 2593-2610.

Marozzi, M., Mukherjee, A. and Kalina, J. (2020), Interpoint distance tests for high-dimensional comparison studies. *J. Appl. Stat.*, **47**(4), 653-665.

Modarres, R. and Song, Y. (2020), Interpoint Distances: Applications, Properties and Visualization. *Applied Stochastic Models in Business and Industry*, https://doi.org/10.1002/asmb.2508

Modarres, R. (2020), Nonparametric Tests for Detection of High Dimensional Outliers. Submitted for publication.

Muirhead R. J. (1982), *Aspects of Multivariate Statistical Theory*, John Wiley & Sons, New York, NY.

Pal, A. K., Mondal, P. K., and Ghosh, A. K. (2016), High dimensional nearest neighbor classification based differences of inter-point distances. *Pattern Recognition Letters*, **74**,1–8.

Robert, P., Cléroux, R., and Ranger, N. (1985), Some results on vector correlation. *Computational Statistics and Data Analysis*, **3**, 25-32.

Sarkar, S. and Ghosh, A. K. (2020), On Perfect Clustering of High Dimension, Low Sample Size Data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(9), 2257-2272.

Song, Y. and Modarres, R. (2019), Interpoint Distance Test of Homogeneity for Multivariate Mixture Models. *International Statistical Review*, **87**(3), 613–638.

Srivastava, M. S. (2005), Some tests concerning the covariance matrix in high-dimensional data. *Journal of Japan Statistical Society*, **35**, 251-272.

Srivastava, M. S. and Kubokawa, T. (2013), Tests for multivariate analysis of variance in high dimension under non-normality. *Journal of Multivariate Analysis*, **115**, 204-216.

Appendix

Proof. **Theorem** 2.1:

Since functions of independent random vectors are independent, the necessary part follows immediately. To prove the sufficient part we need to show that if IPDs $\|\mathbf{X}_1^{(r)} - \mathbf{X}_2^{(r)}\|^2$ and $\|\mathbf{X}_1^{(s)} - \mathbf{X}_2^{(s)}\|^2$ are independent, then $\mathbf{X}^{(r)}$ and $\mathbf{X}^{(s)}$ are independent. Since the contrapositive of any true proposition is also true, we will show if $\mathbf{X}^{(r)}$ and $\mathbf{X}^{(s)}$ are dependent, then $d^2_{(X^{(r)})}$ and $d^2_{(X^{(s)})}$ are dependent.

Suppose $[\mathbf{X}^{(r)}, \mathbf{X}^{(s)}]'$ follows a multivariate normal distribution with mean $(\boldsymbol{\mu}^{(r)}, \boldsymbol{\mu}^{(s)})'$ and covariance matrix $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{rr} & \boldsymbol{\Sigma}_{rs} \\ \boldsymbol{\Sigma}_{sr} & \boldsymbol{\Sigma}_{ss} \end{bmatrix}$. Let $\sigma_{(r)t_1t_2}, \sigma_{(s)u_1u_2}, \sigma_{(rs)tu}$ be the components of $\Sigma_{rr}, \Sigma_{ss}, \Sigma_{rs}$, respectively, for $s, u_1, u_2 = 1, \cdots, p_s;\ r, t_1, t_2 = 1, \cdots, p_r$. Suppose $d^2_{(X^{(r)})}$ and $d^2_{(X^{(s)})}$are dependent, we have

$$tr(\Sigma_{rs}^2) = \sum_{t=1}^{p_r} \sum_{u=1}^{p_s} \sigma^2_{(rs)tu} > 0. \tag{8.1}$$

Let $X_{it}^{(r)}$ be the tth variable of the random vector $\mathbf{X}_i^{(r)}$, and $X_{is}^{(s)}$ be the sth variable of the random vector $\mathbf{X}_i^{(s)}$ for $i = 1, 2,\ t = 1, \cdots, p_r,\ s = 1, \cdots, p_s$. One can show that

$$\begin{aligned}
\mathbb{C}\mathrm{ov}\left(\mathrm{d}^2_{(\mathrm{X}^{(r)})}, \mathrm{d}^2_{(\mathrm{X}^{(s)})}\right) &= \mathbb{C}\mathrm{ov}\left(\sum_{t=1}^{p_r} (\mathrm{X}_{1t}^{(r)} - \mathrm{X}_{2t}^{(r)})^2, \sum_{u=1}^{p_s} (\mathrm{X}_{1u}^{(s)} - \mathrm{X}_{2u}^{(s)})^2\right) \\
&= \sum_{t=1}^{p_r} \sum_{u=1}^{p_s} 4\sigma_{(r)tt}\sigma_{(s)uu} \mathbb{C}\mathrm{ov}\left(\mathrm{W}_{1t}^2, \mathrm{W}_{2u}^2\right),
\end{aligned}$$

where $W_{1t} = \frac{X_{1t}^{(r)} - X_{2t}^{(r)}}{\sqrt{2\sigma_{(r)tt}}} \sim N(0,1)$ and $W_{2u} = \frac{X_{1u}^{(s)} - X_{2u}^{(s)}}{\sqrt{2\sigma_{(s)uu}}} \sim N(0,1)$.
The covariance between W_{1t} and W_{2u} is

$$\begin{aligned}
\mathbb{C}\mathrm{ov}\left(\mathrm{W}_{1t}, \mathrm{W}_{2u}\right) &= \frac{\mathbb{C}\mathrm{ov}\left(\mathrm{X}_{1t}^{(r)} - \mathrm{X}_{2t}^{(r)}, \mathrm{X}_{1u}^{(s)} - \mathrm{X}_{2u}^{(s)}\right)}{4\sigma_{(r)tt}\sigma_{(s)uu}} \\
&= \frac{\mathbb{C}\mathrm{ov}\left(\mathrm{X}_{1t}^{(r)}, \mathrm{X}_{1u}^{(s)}\right) + \mathbb{C}\mathrm{ov}\left(\mathrm{X}_{2t}^{(r)}, \mathrm{X}_{2u}^{(s)}\right)}{4\sigma_{(r)tt}\sigma_{(s)uu}} \\
&= \frac{1}{2}\rho_{(rs)tu},
\end{aligned}$$

where $\rho_{(rsy)tu}$ is the correlation between the t-th component of $X^{(r)}$ and u-th component of $X^{(s)}$ vectors. Thus, the W_{1t} and W_{2u} are jointly normal random variables with $\frac{1}{2}\rho_{(rs)tu}$ correlation. The conditional expectation and variance of W_{1t} given W_{2u} are $E(W_{1t}|W_{2u}) = \rho_{(rs)tu}W_{2u}$, and $\mathrm{Var}(W_{1t}|W_{2u}) = 1 - \rho^2_{(rs)tu}$. Thus, we have $E(W^2_{1t}|W_{2u}) = 1 - \rho^2_{(rs)tu} + \rho^2_{(rs)tu}W^2_{2u}$. The covariance between W^2_{1t} and W^2_{2u} is

$$\begin{aligned}\mathbb{C}\mathrm{ov}(\mathrm{W}^2_{1t}, \mathrm{W}^2_{2u}) &= E(W^2_{1t}W^2_{2u}) - E(W^2_{1t})E(W^2_{2u}) = E(E(W^2_{1t}W^2_{2u}|W_{2u})) - 1\\ &= E[(1 - \rho^2_{(rs)tu} + \rho^2_{(rs)tu}W^2_{2u})W^2_{2u}] - 1\\ &= \rho^2_{(rs)tu}(E(W^4_{2u}) - 1) = 2\rho^2_{(rs)tu}.\end{aligned}$$

Since $\mathbb{C}\mathrm{ov}\left(\mathrm{W}^2_{1t}, \mathrm{W}^2_{2u}\right) = 2\rho^2_{(rs)tu}$, we obtain

$$\begin{aligned}\mathbb{C}\mathrm{ov}\left(\mathrm{d}^2_{(\mathrm{X}^{(r)})}, \mathrm{d}^2_{(\mathrm{X}^{(s)})}\right) &= \sum_{t=1}^{p_r}\sum_{u=1}^{p_s} 4\sigma_{(r)tt}\sigma_{(s)uu}\mathbb{C}\mathrm{ov}\left(\mathrm{W}^2_{1t}, \mathrm{W}^2_{2u}\right)\\ &= \sum_{t=1}^{p_r}\sum_{u=1}^{p_s} 8\sigma_{(r)tt}\sigma_{(s)uu}\rho^2_{(rs)tu}\\ &= 8\sum_{t=1}^{p_r}\sum_{u=1}^{p_s} \mathbb{C}\mathrm{ov}^2(\mathrm{X}^{(r)}_{1t}, \mathrm{X}^{(s)}_{1u}) = 8\mathrm{tr}(\Sigma'_{rs}\Sigma_{rs}).\end{aligned}$$

Under the assumption of dependence, the above covariance and the equation (8.1) prove that $\mathbb{C}\mathrm{ov}(\mathrm{d}^2_{(\mathrm{X}^{(r)})}, \mathrm{d}^2_{(\mathrm{X}^{(s)})}) > 0$. Thus, $d^2_{(X^{(r)})}$ and $d^2_{(X^{(s)})}$ are dependent.

Let ρ_x and ρ_y denote the coefficient of correlations in the bivariate power normal and bivariate normal scales, respectively. Freeman and Modarres (2006) study the properties of the power normal distribution and show that

$$\rho_x = h(\rho_y) = \frac{\sum_{i=1}^{\infty} i! b_{1i} b_{2i}\, \rho_y^i}{\sqrt{(\sum_{i=1}^{\infty} b^2_{1i} i!)(\sum_{j=1}^{\infty} b^2_{2j} j!)}},$$

where the form of the function h depends on the mean, covariance and the transformation parameter. Here, b_{1i}, b_{2i} are defined in terms of the parameters and the Chebyshev-Hermite polynomials. It follows that $\rho_y = 0$ if and only if $\rho_x = 0$ and based on the first part of the Theorem, the proof is completed. □

Proof. **Lemma** 3.1:

Let $C = \mathbf{\Sigma}^{1/2}$ be the square root of $\mathbf{\Sigma}$. One can write $\mathbf{Z} = \mathbf{\Sigma}^{-1/2}\left(\mathbf{X} - \boldsymbol{\mu}\right)$. Let $\mathbf{P}$ be a $p \times p$ orthogonal matrix which diagonalizes $\mathbf{\Sigma}^{1/2} A \mathbf{\Sigma}^{1/2}$, where $\mathbf{PP'} = \mathbf{P'P} = \mathbf{I}_p$. That is, $\mathbf{P'\Sigma^{1/2}A\Sigma^{1/2}P} = \mathbf{\Lambda}$, where $\mathbf{\Lambda} = diag(\lambda_1, \ldots, \lambda_p)$, and $\lambda_1, \ldots, \lambda_p$ are the eigenvalues of $\mathbf{P'\Sigma^{1/2}A\Sigma^{1/2}P}$. Let $\mathbf{U} = \boldsymbol{P'Z}$, and note that $\mathbf{Z} = \mathbf{PU}$, $\mathbb{E}(\mathbf{U}) = \mathbf{0}$ and $\mathbb{C}\mathrm{ov}(\mathbf{U}) = \mathbf{I}_p$ because $\mathbb{E}(\mathbf{Z}) = \mathbf{0}$ and $\mathbb{C}\mathrm{ov}(\mathbf{Z}) = \mathbf{I}_p$ under the $\mathcal{M}$ model. We obtain the quadratic form (3.1) since

$$\begin{aligned}
Q(\mathbf{X}) = \mathbf{X'AX} &= (\mathbf{\Sigma}^{1/2}\mathbf{Z} + \boldsymbol{\mu})'\mathbf{A}(\mathbf{\Sigma}^{1/2}\mathbf{Z} + \boldsymbol{\mu}) \\
&= (\mathbf{Z} + \mathbf{\Sigma}^{-1/2}\boldsymbol{\mu})'(\mathbf{\Sigma}^{1/2}A\mathbf{\Sigma}^{1/2})(\mathbf{Z} + \mathbf{\Sigma}^{-1/2}\boldsymbol{\mu}) \\
&= (\mathbf{U} + \mathbf{b})'\mathbf{P'\Sigma^{1/2}A\Sigma^{1/2}P}(\mathbf{U} + \mathbf{b}) \\
&= (\mathbf{U} + \mathbf{b})'\mathbf{\Lambda}(\mathbf{U} + \mathbf{b}) = \begin{cases} \sum_{t=1}^{p} \lambda_t (U_t + b_t)^2, & \text{if } \boldsymbol{\mu} \neq \mathbf{0}, \\ \sum_{t=1}^{p} \lambda_t U_t^2, & \text{if } \boldsymbol{\mu} = \mathbf{0}, \end{cases}
\end{aligned}$$

where $\mathbf{b}' = (\mathbf{P'\Sigma}^{-1/2}\boldsymbol{\mu})' = (b_1, \ldots, b_p)$ and $\mathbf{U}' = (U_1, \ldots, U_p)$. □

Proof. **Lemma** 3.2:

Let $\mathbf{X}_1$ and $\mathbf{X}_2$ be i.i.d. vectors drawn from an elliptical distribution $E_p(\mu, \mathbf{\Sigma}, \psi)$. Consider the distribution of IPD $d^2(\mathbf{X}_1, \mathbf{X}_2)$. Note that $\mathbf{X}_i = \boldsymbol{\mu} + \mathbf{BT}_i$ where $\mathbf{T}_i \sim \mathbb{S}_p(\psi)$, for $i = 1, 2$, and $\mathbf{B}$ is a $p \times p$ matrix such that $\mathbf{\Sigma} = \mathbf{BB'}$ is positive definite. It follows that $\mathbf{X}_1 - \mathbf{X}_2 = B(\mathbf{T}_1 - \mathbf{T}_2) \sim E_p(\mathbf{0}, 2\mathbf{\Sigma}, \psi)$. Hence, $d^2\left(\mathbf{X}_i, \mathbf{X}_j\right) = (\mathbf{T}_1 - \mathbf{T}_2)'\mathbf{\Sigma}(\mathbf{T}_1 - \mathbf{T}_2) = \mathbf{R'\Sigma R}$ where $\mathbf{R} = \sqrt{2}(\mathbf{T}_1 - \mathbf{T}_2) \sim \mathbb{S}_p(\psi)$. Let $diag(\lambda_1, \ldots, \lambda_p)$ be the eigenvalues of $2\mathbf{\Sigma}$. It follows that $d^2(\mathbf{X}_1, \mathbf{X}_2) = \mathbf{R'\Sigma R} = \mathbf{R}'diag(\lambda_1, \ldots, \lambda_p)\mathbf{R} = \sum_{j=1}^{p} \lambda_j U_j^2$ where U_j has a univariate symmetric distribution with center zero. □

Proof. **Lemma** 4.1:

For $t \in \mathbf{R}$ and $r = 1, \ldots, k$, using (4.4), one can show that $[tr(\mathbf{S})]^t = \frac{1}{2^t}\bar{d}_p^{\,2t}$ and $[tr(\mathbf{S}_{rr})]^t = \frac{1}{2^t}\bar{d}_{p_r}^{\,2t}$. Let $\lambda_1, \ldots, \lambda_p$ denote the eigenvalues of $\mathbf{S}$. It follows from the Cauchy-Schwarz inequality that $\left(\sum_{i=1}^{p} \lambda_i^t\right)^2 \leq p\left(\sum_{i=1}^{p} \lambda_i^{2t}\right)$ with equality holding if and only if $\lambda_1 = \lambda_2 = \ldots, \lambda_p = \lambda$, for some constant λ. Since $tr\left(\mathbf{S}^t\right) = \sum_{i=1}^{p} \lambda_i^t$ for $t \geq 1$ it follows that $\left(tr\mathbf{S}\right)^2 \leq p\ tr\left(\mathbf{S}^2\right)$ when $t = 1$. Coupled with (4.4) allows one to show that $\bar{d}_p^{\,2} \leq_{St} 2\sqrt{p\ tr(\mathbf{S}^2)}$ and $\bar{d}_{p_r}^{\,2} \leq_{St} 2\sqrt{p_r\ tr(\mathbf{S}_{rr}^2)}$ where $\leq_{St}$ is the stochastic ordering between two random variables. □