High-dimensional CLTs for individual Mahalanobis distances

Holgersson T.
Dai D.

May, 2014

High-dimensional CLTs for individual Mahalanobis distances

Holgersson. T. [1,2] Dai. D. [1]

Abstract

In this paper we derive central limit theorems for two different types of Mahalanobis distances in situations where the dimension of the parent variable increases proportionally with the sample size. It is shown that although the two estimators are closely related and behave similarly in finite dimensions, they have different convergence rates and are also centred at two different points in high-dimensional settings. The limiting distributions are shown to be valid under some general moment conditions and hence available in a wide range of applications.

Keywords: Mahalanobis distance, increasing dimension, weak convergence, Marcenko-Pastur distribution, outliers, Pearson family.

1. Introduction

The Mahalanobis distance (MD) (Mahalanobis, 1936; Anderson, 2003; Mardia et al., 1980) appears in various statistical contexts such as in outlier detection, cluster analysis or in assessment of distributional properties of random vectors. While

---

[1]Linnaeus university, SE 351 95, Växjö, Sweden
[2]Jönköping university, SE 551 11, Jönköping, Sweden

the properties of the sample MD are fairly well established in finite dimensions (i.e. when the dimension p of the random vector of interest is a fixed number), some care needs to be taken in cases when p increases together with the sample size n such that as $p/n \rightarrow c\,(0,1)$ as $n \rightarrow \infty, p \rightarrow \infty$. Some important matters were discussed in Dai, Holgersson and Karlsson (2014), where it was shown that the standard MD estimator has a too low range relative to the true unobservable MD and that the leave-one-out MD estimator, derived by excluding a specific observation from the estimation of the mean vector and covariance matrix on which the MD depends, is asymptotically biased. In this paper we derive some asymptotic distributions of Mahalanobis distances in high-dimensional settings as defined above, since such results do not appear to be available in the literature. Limiting distributions will in turn be important for making inference of, for example, a potential single outlier or other inferences requiring a distribution. Specifically, we apply recent findings of limiting spectral distributions (Bai and Silverstein, 2009; Bai et al., 2007) which apply directly to the sample MD.

2. Mahalanobis distances

This paper is concerned with the MD defined by the scaled distance between an individual observation and its expected value, sometimes referred to as individual Mahalanobis distances. Formally, we make the following definition:

Definition 1. *Let* $\mathbf{X}_i : p \times 1$ *be a random vector such that* $E\,[\mathbf{X}_i] = \boldsymbol{\mu}_{p \times 1}$ *and* $E\,[(\mathbf{X}_i - \boldsymbol{\mu})\,(\mathbf{X}_i - \boldsymbol{\mu})'] = \boldsymbol{\Sigma}_{p \times p}$ *for* $i = 1, ..., n$. *Then we make the following definition:*

$$D_{ii} := (\mathbf{X}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}\,(\mathbf{X}_i - \boldsymbol{\mu})\,. \tag{1}$$

The D_{ii} statistic measures the scaled distance between an individual observation $\mathbf{X}_i$ and its expected value $\boldsymbol{\mu}$, which is frequently used to display data, assess distributional properties and detect influential values. Estimators of (1) may be obtained by simply replacing the unknown parameters with appropriate estimators. In case both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown and replaced by the standard estimators we get the well-known estimators defined below:

Definition 2. *Let* $\bar{\mathbf{X}} = n^{-1}\mathbf{X}'\mathbf{1}$ *and* $\mathbf{S} = n^{-1}\sum_{i=1}^{n}\left(\mathbf{X} - \bar{\mathbf{X}}\right)\left(\mathbf{X} - \bar{\mathbf{X}}\right)'$. *Then we make the following definition:*

$$d_{ii} := \left(\mathbf{X}_i - \bar{\mathbf{X}}\right)'\mathbf{S}^{-1}\left(\mathbf{X}_i - \bar{\mathbf{X}}\right). \tag{2}$$

Further discussion of the statistic (2) is available in Mardia (1977) and Mardia et al. (1980) Dai, Holgersson, Karlsson (2014). Note that although the inverse sample covariance matrix is sometimes expressed using a different divisor (for example *n-p-1* instead of *n*), the statistic (2) is unbiased in the sense that $E\left[d_{ii}\right] = E\left[D_{ii}\right] = p$.

Alternative estimators are available through leave-one-out estimators, obtained by omitting a specific observation from the estimation of the sample mean vector and the sample covariance matrix, and thereby avoiding correlations between the components within the MD estimator. Formally, this is done as follows:

Definition 3. *Let* $\bar{\mathbf{X}}_{(i)} := (n-1)^{-1}\sum_{k=1,k\neq i}^{n}\mathbf{X}_k$ *and* $\mathbf{S}_{(i)} :=$ $(n-1)^{-1}\sum_{k=1,k\neq i}^{n}\left(\mathbf{X}_k - \bar{\mathbf{X}}_{(i)}\right)\left(\mathbf{X}_k - \bar{\mathbf{X}}_{(i)}\right)'$. *Then the following alternative estimator of* (1) *is defined:*

$$d_{(ii)} := \left(\mathbf{X}_i - \bar{\mathbf{X}}_{(i)}\right)'\mathbf{S}_{(i)}^{-1}\left(\mathbf{X}_i - \bar{\mathbf{X}}_{(i)}\right). \tag{3}$$

Note that in case of independently identically normally distributed data the sample mean vector and the sample covariance matrix are independent and hence all components within the estimator (3) will be mutually independent, which results in estimators whose distributional properties differ from that of (2) and are also robust to a single outlier in the sense of not contaminating the estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}^{-1}$. They will hence be included for further investigations in the paper.

The MD distances has mainly been used in empirical works and very few results are available about their behavior in high-dimensional settings, i.e. in cases where the sample size n is proportional to the dimension of the random vector p. The reminder of this section is therefore concerned with the asymptotic distributions of (2) and (3) under such settings.

Proposition 1. *Let* $\mathbf{X}_i \overset{iid}{\sim} N\left(\mu_{p \times 1}, \boldsymbol{\Sigma}_{p \times p}\right)$ *where* $\sup_l \{\lambda_l\}_{l=1}^p \leq r < \infty$ *and* $\inf_l \{\lambda_l\}_{l=1}^p \geq r' > 0$ *where* λ_l *are the eigenvalues of* $\boldsymbol{\Sigma}_{p \times p}$. *Moreover, suppose we have a sample* $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$ *such that* $p/n \to c$ *where* $0 \leq c < 1$. *Then,*

$$\frac{\sqrt{p}}{\sqrt{2}}\left(p^{-1}D_{ii} - 1\right) \overset{\ell}{\to} N(0,1) \text{ as } n \to \infty, p \to \infty. \qquad (4)$$

Proof. Since D_{ii} is invariant to linear transformations of the type $\mathbf{X}_i \mapsto \mathbf{A} + \mathbf{B}\mathbf{X}_i$ we may assume $\mathbf{X}_i\left(\mathbf{0}, \mathbf{I}\right)$. Hence $D_{ii} \sim \chi^2_{(p)}$ and Proposition 1 follows. ∎

The asymptotic distribution obtained in Proposition 1 may easily be derived under some general moment restrictions of the parent variable $\mathbf{X}_i$ and relax the normality assumption. In real applications, however, the (inverse) covariance matrix is unknown and has to be estimated, a problem which in turn bares on the distribution of $\mathbf{X}_i$. We will keep the normality assumption of Proposition 1 in the derivation of the asymptotic properties of sample Mahalanobis distances in order to access known finite-sample properties, and later on relax this assumption.

Theorem 1. *Let* $\mathbf{X}_i$ *be distributed as in Proposition 1. Then,*

$$\frac{\sqrt{p}}{\sqrt{2}} \frac{(p^{-1}d_{ii} - 1)}{\sqrt{(1-c)}} \xrightarrow{\ell} N(0,1) \text{ as } n \to \infty, p \to \infty. \tag{5}$$

Proof. Let $\mathbf{G} = (d_{ij})$ be the matrix of sample Mahalanobis distances as defined in Definition 2 and let $\mathbf{H}_p = \mathbf{I} - n^{-1}\mathbf{1}\mathbf{1}'$ where $\mathbf{1}$ is a vector of ones of appropriate dimension. Following Mardia (1977) we have that

$$n^{-1}\mathbf{H}_r^{-1/2}\mathbf{G}_r\mathbf{H}_r^{-1/2} \sim \mathbf{B}_r\left(\frac{1}{2}p, \frac{1}{2}(n-p-1)\right), \tag{6}$$

where $\mathbf{G}_r$ and $\mathbf{H}_r$ are principal submatrixes of order r from $\mathbf{G}$ and $\mathbf{H}$ respectively and $\mathbf{B}_r$ has a matrix $r \times r$ Type I Beta distribution. We then have that $n^{-1}\mathbf{H}_1^{-1/2}\mathbf{G}_1\mathbf{H}_1^{-1/2} = n^{-1}(1-n^{-1})^{-1/2}g_{11}(1-n^{-1})^{-1/2} = g_{11}/(n-1) = M_{11}$, say, which is distributed as a univariate type-I Beta distribution. Specifically, $M_{ii} \sim B_I(\alpha, \beta)$ where $\alpha = p/2$, $\beta = (n-p-1)/2$. A convenient re-parameterization (Mielke Jr, 1975; Balding and Nichols, 1995) is given by setting $\nu = \alpha + \beta = (n-1)/2$ and $\mu = \alpha/(\alpha+\beta) = p/(n-1)$ which yields the moments,

$$\mu = E[M_{ii}], \tag{7}$$

$$\sigma^2 = V[M_{ii}] = \mu(1-\mu)/(1+\nu), \tag{8}$$

$$\gamma_1 = \frac{E[M_{ii} - \mu]^3}{(V[M_{ii}])^{3/2}} = \frac{2(1-2\mu)\sqrt{1+\nu}}{(2+\nu)\sqrt{\mu(1-\mu)}}, \tag{9}$$

$$\gamma_2 = \frac{E[M_{ii} - \mu]^4}{(V[M_{ii}])^2} - 3 = \frac{6}{3+\nu}\left(\frac{(1-2\mu)^2(1+\nu)}{\mu(1-\mu)(2+\nu)} - 1\right) \tag{10}$$

In other words the mean, variance, skewness and excess kurtosis of M_{ii} are given by μ, σ^2, γ_1 and γ_2 respectively as defined above. Substituting n and p into μ, σ^2, γ_1 and γ_2 and taking limits it is found that $\lim_{n,p\to\infty} \mu = c$, $\lim_{n,p\to\infty} \sigma^2 = 2c(1-c)/n$, $\lim_{n,p\to\infty} \gamma_1 = 0$, $\lim_{n,p\to\infty} \gamma_2 = 0$ and hence the beta distributed M_{ii} limits a normal distribution in the Pearson chart (Ord, 1972) determined by the plane of γ_1^2 and γ_2^2. Finally, noting that γ_1 and γ_2 are invariant to linear transformations of the

5

type $M_{ii} \mapsto a + bM_{ii}$ (so that centering and scaling does not affect the skewness or kurtosis coefficients) the proof is complete.

■

Remark 1: The assumption $\inf_l \{\lambda_l\}_{l=1}^p \geq r' > 0$ of theorem 3 and 4 is necessary to ensure that Σ is of full rank, which in turn (assuming $p < n + 1$) implies full rank of $\mathbf{S}$ and hence existence of the unique inverse $\mathbf{S}^{-1}$. The second assumption, $\sup_l \{\lambda_l\}_{l=1}^p \leq r < \infty$, appears to be necessary due to the eigenvalue relation $\lambda_{(j)}(\Sigma^{-1}) = (\lambda_{(j)}(\Sigma))^{-1}$ where $\lambda_{(j)}(\mathbf{A})$ denotes the jth ordered eigenvalue of $\mathbf{A}$ since, if the largest eigenvalue λ_p is allowed to increase unboundedly as $p \to \infty$, then $\lambda_{(p)}(\Sigma^{-1})$ limits zero in which case $\mathbf{S}^{-1}$ will limit a singular matrix which in turn means that the sample MD is not uniquely defined.

Remark 2: Note that when $c = 1$ the distribution in Theorem 1 becomes degenerate since when $n = p$ the statistic $\mathbf{S}^{-1}$ does not exist and hence d_{ii} is un-defined.

Below we will drive a CLT similar to that of Theorem 1 for the leave-one-out MD, which limits a different point than the standard MD and hence requires a different formulation. Specifically, we have the following:

Theorem 2. *Let* $\mathbf{X}_i$ *be distributed as in Proposition 1. Then,*
$$\frac{\sqrt{p}}{\sqrt{2}}\sqrt{(1-c)}\left(p^{-1}(1-c)d_{(ii)} - 1\right) \overset{\ell}{\to} N(0,1) \text{ as } n \to \infty, p \to \infty.$$

Proof. Firstly, we note that d_{ii} is invariant to linear transformations $\mathbf{X}_i \mapsto \mu + \Sigma^{1/2}\mathbf{X}_i$ for any positive semi-definite Σ. Hence we may assume that $\mathbf{X}_i \overset{iid}{\sim} N(\mathbf{0}, \mathbf{I})$ and therefore $\left(\mathbf{X}_i - \bar{\mathbf{X}}_{(i)}\right) \sim N\left(\mathbf{0}, \left(\frac{n}{n-1}\right)\mathbf{I}\right)$.

6

Moreover, $(n-2)\,\mathbf{S}_{(i)} \sim Wishart\,(n-2,\mathbf{I})$ (Bilodeau and Brenner, 1999). Noting that $\left(\dfrac{n-1}{n}\right)^{1/2}\left(\mathbf{X}_i - \bar{\mathbf{X}}_{(i)}\right) \sim N\left(\mathbf{0},\mathbf{I}\right)$ it follows that $d_{(ii)} = (n-2)^{-1}\left(\dfrac{n-1}{n}\right)\left(\mathbf{X}_i - \bar{\mathbf{X}}_{(i)}\right)'\mathbf{S}_{(ii)}^{-1}\left(\mathbf{X}_i - \bar{\mathbf{X}}_{(i)}\right) \sim \left(\dfrac{(n-2)\,p}{n-p-1}\right)F_{(p,n-p-1)}$ (Mardia et al., 1980). Hence, $p^{-1}d_{(ii)} \sim \left(\dfrac{n-2}{n-p-1}\right)F_{(p,n-p-1)} \xrightarrow{\ell} \left(\dfrac{n}{n-p}\right)F_{(p,n-p)} = \left(\dfrac{1}{1-c}\right)F_{(p,p(c^{-1}-1))}$ or $(1-c)\,p^{-1}d_{(ii)} \xrightarrow{\ell} F_{(p,p(c^{-1}-1))}$.

The expected value, variance, skewness and excess kurtosis of this distribution are given Johnson et al. (1995) below:

$$\mu = E\left[F_{(d_1,d_2)}\right] = \frac{d_2}{d_2 - 2},$$
$$\sigma^2 = V\left[F_{(d_1,d_2)}\right] = \frac{2d_2^2\,(d_1 + d_2 - 2)}{d_1(d_2 - 2)^2\,(d_2 - 4)},$$
$$\gamma_1 = \frac{(2d_1 + d_2 - 2)\sqrt{2\,(d_2 - 4)}}{(d_2 - 6)\sqrt{d_1\,(d_1 + d_2 - 2)}},$$
$$\gamma_2 = 12\frac{d_1\,(5d_2 - 22)\,(d_1 + d_2 - 2) + (d_2 - 4)\,(d_2 - 2)^2}{d_1\,(d_2 - 6)\,(d_2 - 8)\,(d_1 + d_2 - 2)}.$$

where $d_1 = p$ and $d_2 = n - p - 1$. Therefore, $\lim\limits_{n,p\to\infty} E\left[(1-c)\,p^{-1}d_{(ii)}\right] = 1$ and $\lim\limits_{n,p\to\infty} V\left[(1-c)\,p^{-1}d_{(ii)}\right] = 2/p\,(1-c)$. The limits $\lim\limits_{n,p\to\infty}\gamma_1 = 0$, $\lim\limits_{n,p\to\infty}\gamma_2 = 0$ are established in Appendix. It is well known (Ord, 1972) that the F-distribution is a particular example of the Beta prime distribution, which belongs to the Pearson type VI distribution. And since the skewness and excess kurtosis both limits zero, the standardized $d_{(ii)}$ asymptotically converges to a normal distribution and Theorem 2 is established.

■

Although Theorems 1 and 2 above assume a Gaussian distribution of the parent variable $\mathbf{X}$, one may expect a limiting Gaussian distribution to be valid under more general distributional properties of $\mathbf{X}$. Below we will show that this is indeed the case. Firstly, we will state a useful identity:

Remark 3: Duality between d_{ii} and $d_{(ii)}$:

Set $\mathbf{W} = n\mathbf{S}$ and $\mathbf{W}_{(i)} = (n-1)\mathbf{S}_{(i)}$. Then we have that $\left|\mathbf{W}_{(i)} + \mathbf{Y}_i\mathbf{Y}'_i\right| = \left|\mathbf{W}_{(i)}\right|\left(1 + \mathbf{Y}'_i\mathbf{W}_{(i)}^{-1}\mathbf{Y}\right)$ and $\left|\mathbf{W} - \mathbf{Y}_i\mathbf{Y}'_i\right| = |\mathbf{W}|\left(1 - \mathbf{Y}'_i\mathbf{W}^{-1}\mathbf{Y}\right)$ and so $\dfrac{\left|\mathbf{W}_{(i)} + \mathbf{Y}_i\mathbf{Y}'_i\right|}{\left|\mathbf{W}_{(i)}\right|} = 1 + n^{-1}d_{(ii)}$ and $\dfrac{|\mathbf{W} - \mathbf{Y}_i\mathbf{Y}'_i|}{|\mathbf{W}|} = 1 - n^{-1}d_{ii}$.

Hence $\left(1 - cp^{-1}d_{ii}\right)\left(1 + cp^{-1}d_{(ii)}\right) = 1$ from which we get the identities $p^{-1}d_{ii} = \dfrac{p^{-1}d_{(ii)}}{1 + cp^{-1}d_{(ii)}}$ and $p^{-1}d_{(ii)} = \dfrac{p^{-1}d_{ii}}{1 - cp^{-1}d_{ii}}$. Moreover, since $1 \leq \left(1 + cp^{-1}d_{(ii)}\right)$ it follows that $0 \leq cp^{-1}d_{ii} \leq 1$, or equivalently $0 \leq p^{-1}d_{ii} \leq c^{-1}$, which explicitly shows how the range of $p^{-1}d_{ii}$ is restricted by c. In particular, $p^{-1}d_{ii}$ will have a too low range relative to $p^{-1}D_{ii}$ for large values of c (for example, if $\mathbf{X}_i$ is normally distributed then $p^{-1}d_{(ii)} \sim p^{-1}\chi^2_{(p)}$, which is not bounded above).

Theorem 3. *Let $\mathbf{X}_i$ be iid distributed with $E[(X_{ij} - \mu_j)/\sigma_j]^4 \leq \nu < \infty, j = 1, ..., p, i = 1, ..., n$ and $p/n \to c$ where $0 < c < 1$. Then,*
$$\frac{\sqrt{p}}{\sqrt{2}}\sqrt{(1-c)}\left((1-c)\,p^{-1}d_{(ii)} - 1\right) \xrightarrow{\ell} N(0,1) \text{ as } n \to \infty, p \to \infty$$

Proof. From [Jonsson (1982)](); [Bai et al. (2007)](), we have that $\frac{\sqrt{p}}{\sqrt{2}}\left(p^{-1}tr\left(\tilde{\mathbf{S}}_{(i)}^{-1}\right) - (1-c)^{-1}\right)$ converges weakly to a normal distribution with zero mean and variance $\int \frac{1}{z^2}dF_c(z) - \left(\int \frac{1}{z}dF_c(z)\right)^2$ where $dF_c(z) = \frac{1}{2\pi zc}\sqrt{\left((1 + \sqrt{c})^2 - z\right)\left(z - (1 - \sqrt{c})^2\right)}$ is the standard Marcenko-Pastur distribution ([Marchenko and Pastur, 1967](); [Bai]()

et al., 2007). It may be shown that $\int \frac{1}{z} dF_c(z) = (1-c)^{-1}$ and that $\int \frac{1}{z^2} dF_c(z) = \frac{1}{(1-c)^3}$ (Arharov, 1971; Serdobolskii, 2000; Glombek, 2014) so that

$$V\left[\frac{\sqrt{p}}{\sqrt{2}}\left((1-c)\,p^{-1}d_{(ii)} - 1\right)\right] \to \left(\frac{1}{1-c}\right) \text{ which completes the proof.}$$

$\blacksquare$

Theorem 4. *Let $\mathbf{X}_i$ be distributed as in Theorem 3. Then*
$$\frac{\sqrt{p}}{\sqrt{2}} \frac{(p^{-1}d_{ii} - 1)}{\sqrt{(1-c)}} \xrightarrow{\ell} N(0,1) \text{ as } n \to \infty, p \to \infty$$

Proof. From Theorem 3, we know that $\frac{\sqrt{p}}{\sqrt{2}}(1-c)\,p^{-1}d_{(ii)} \xrightarrow{\ell} N\left(\frac{\sqrt{p}}{\sqrt{2}}, \frac{1}{1-c}\right)$ as $n \to \infty, p \to \infty$ or $\frac{\sqrt{p}}{\sqrt{2}}p^{-1}d_{(ii)} \xrightarrow{\ell} N\left(\frac{\sqrt{p}}{\sqrt{2}}\frac{1}{1-c}, \frac{1}{(1-c)^3}\right)$. Since $\frac{\sqrt{p}}{\sqrt{2}}p^{-1}d_{ii} = \frac{\sqrt{p}}{\sqrt{2}}\frac{p^{-1}d_{(ii)}}{1 + cp^{-1}d_{(ii)}}$ is a simple measurable function we can apply Cramer's theorem (Ferguson, 1996) to get $\left(g\left(p^{-1}d_{(ii)}\right) - g(\mu)\right) \xrightarrow{\ell} N\left(0, \dot{g}^2(\mu)\frac{1}{(1-c)^3}\right)$ where $\mu = \frac{\sqrt{p}}{\sqrt{2}}\frac{1}{1-c}$ and

$$g\left(\frac{\sqrt{p}}{\sqrt{2}}p^{-1}d_{(ii)}\right) = \frac{\frac{\sqrt{p}}{\sqrt{2}}p^{-1}d_{(ii)}}{1 + c\frac{\sqrt{p}}{\sqrt{2}}p^{-1}d_{(ii)}} \text{ is a function of } \frac{\sqrt{p}}{\sqrt{2}}p^{-1}d_{(ii)} \text{ so that } g(\mu) =$$

$$\frac{\mu}{1 + c\frac{\sqrt{2}}{\sqrt{p}}\mu} = g\left(\frac{\sqrt{p}}{\sqrt{2}}\frac{1}{1-c}\right) = \frac{\frac{\sqrt{p}}{\sqrt{2}}\frac{1}{1-c}}{1 + c\frac{\sqrt{p}}{\sqrt{2}}\frac{\sqrt{2}}{\sqrt{p}}\frac{1}{1-c}} = \frac{\sqrt{p}}{\sqrt{2}}\frac{\frac{1}{1-c}}{1 + c\frac{1}{1-c}} = \frac{\sqrt{p}}{\sqrt{2}}. \text{ Then}$$

we find that $\dot{g}^2(\mu) = (\partial g(\mu)/\partial \mu)^2 = \left[\partial\left(\frac{\mu}{1 + c\mu}\right)\Big/\partial\mu\right]^2 = \left(1\Big/\left(1 + c\frac{\sqrt{2}}{\sqrt{p}}\mu\right)^2\right)^2$,

or in terms of $\mu = \frac{\sqrt{p}}{\sqrt{2}}\frac{1}{1-c}$, $\dot{g}^2\left(\frac{1}{1-c}\right) = \left(1\Big/\left(1 + \frac{\sqrt{2}}{\sqrt{p}}c\frac{\sqrt{p}}{\sqrt{2}}\frac{1}{1-c}\right)^2\right)^2 =$

$1\Big/\left(1 + \frac{\sqrt{2}}{\sqrt{p}}c\frac{\sqrt{p}}{\sqrt{2}}\frac{1}{1-c}\right)^4 = (1-c)^4$. Therefore we get that

$\frac{\sqrt{p}}{\sqrt{2}}\frac{(p^{-1}d_{ii} - 1)}{\sqrt{(1-c)}} \xrightarrow{\ell} N(0,1)$ as $n \to \infty, p \to \infty$. $\blacksquare$

Remark 4: Note that, although Theorems 3 and 4 are more general in the sense that they don't require a normal distribution of the parent variable $\mathbf{X}_i$, or even existence of all moments, Theorems 1 and 2 are still relevant since they say a little bit more than Theorems 3-4, namely that the normalized MD distribution limits the normal distribution within the Pearson (γ_1, γ_2) plane (i.e. within the Pearson family of distributions) and hence possess all Pearson family properties for any pairing (n, p) such that $p < n + 1$.

3. summary

In this paper central limit theorems are derived for two types of individual Mahalanobis distances in cases where the dimension of data increases proportionally with the sample size. It is shown that the type of MD estimator obtained by omitting a single observation from the estimation results in an asymptotic distribution different to that of the standard estimator. Moreover, an explicit link between the leave-one-out estimator and the standard estimator is given which in turn can be used to derive various properties of the MD estimator. The limiting distributions derived in the paper only require a finite fourth order moment restriction and will be useful for making inference of a single outlier or other applications requiring a sampling distribution. While some type of inferences, for example of linear combinations or vector-valued statistics of Mahalanobis distances, may be of relevance such problems will be challenging in several senses and is left for future research.

Appendices

A.

$$\gamma_1 = \frac{(2d_1 + d_2 - 2)\sqrt{2(d_2 - 4)}}{(d_2 - 6)\sqrt{d_1(d_1 + d_2 - 2)}} = \frac{(2p + n - p - 1 - 2)\sqrt{2(n - p - 1 - 4)}}{(n - p - 1 - 6)\sqrt{p(p + n - p - 1 - 2)}}$$

$$= \frac{(n + p - 3)\sqrt{2(n - p - 5)}}{(n - p - 7)\sqrt{p(n - 3)}} = \frac{\left(1 + \frac{p}{n} - \frac{3}{n}\right)\sqrt{2\left(1 - \frac{p}{n} - \frac{5}{n}\right)}}{\left(1 - \frac{p}{n} - \frac{7}{n}\right)\sqrt{p\left(1 - \frac{3}{n}\right)}}.$$

Hence,

$$\lim_{n,p \to \infty} \gamma_1 = \lim_{n,p \to \infty} \frac{(1 + c)\sqrt{2(1 - c)}}{(1 - c)\sqrt{cn}} \to \frac{1}{\sqrt{n}} \frac{(1 + c)\sqrt{2(1 - c)}}{(1 - c)} \to 0$$

Also,

$$\gamma_2 = 12\frac{d_1(5d_2 - 22)(d_1 + d_2 - 2) + (d_2 - 4)(d_2 - 2)^2}{d_1(d_2 - 6)(d_2 - 8)(d_1 + d_2 - 2)}$$

$$= 12\frac{p(5[n - p - 1] - 22)(p + n - p - 1 - 2) + (n - p - 1 - 4)(n - p - 1 - 2)^2}{d_1(n - p - 1 - 6)(n - p - 1 - 8)(p + n - p - 1 - 2)}$$

$$= 12\frac{p(5n - 5p - 5 - 22)(n - 3) + (n - p - 5)(n - p - 3)^2}{p(n - p - 7)(n - p - 9)(n - 3)}.$$

Hence,

$$\lim_{n,p \to \infty} \gamma_2 = \lim_{n,p \to \infty} 12\frac{5p(n - p)n + (n - p)(n - p)^2}{p(n - p)(n - p)n}$$

$$\to 12\frac{5np(n - p) + (n - p)^3}{np(n - p)^2}$$

$$\to 12\frac{5p/n + (1 - p/n)^2}{p(1 - p/n)}$$

$$\to 12\frac{5c + (1 - c)^2}{p(1 - c)}$$

$$\to 0.$$

Anderson, T. (2003). An introduction to multivariate statistical analysis.

Arharov, L. (1971). Limit theorems]: or the characteristic roots o]: a sample covariance matrix, *Soviet Math. Dokl*, Vol. 12, pp. 1206–1209.

Bai, Z., Miao, B., Pan, G. et al. (2007). On asymptotics of eigenvectors of large sample covariance matrix, *The Annals of Probability* **35**(4): 1532–1572.

Bai, Z. and Silverstein, J. W. (2009). *Spectral analysis of large dimensional random matrices*, Springer.

Balding, D. J. and Nichols, R. A. (1995). A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity, *Human Identification: The Use of DNA Markers*, Springer, pp. 3–12.

Bilodeau, M. and Brenner, D. (1999). *Theory of multivariate statistics*, Springer.

Ferguson, T. S. (1996). *A course in large sample theory*, Vol. 38, CRC Press.

Glombek, K. (2014). Statistical inference for high-dimensional global minimum variance portfolios, *Scandinavian Journal of Statistics* .

Johnson, N., Kotz, S. and Balakrishnan, N. (1995). *Continuous univariate distributions*, number çňň 2 åĲů in *Wiley series in probability and mathematical statistics: Applied probability and statistics*, Wiley & Sons.
URL: *http://books.google.se/books?id=0QzvAAAAMAAJ*

Jonsson, D. (1982). Some limit theorems for the eigenvalues of a sample covariance matrix, *Journal of Multivariate Analysis* **12**(1): 1–38.

Mahalanobis, P. (1936). On the generalized distance in statistics, *Proceedings of the National Institute of Sciences of India*, Vol. 2, New Delhi, pp. 49–55.

Marchenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices, *Matematicheskii Sbornik* **114**(4): 507–536.

Mardia, K. (1977). Mahalanobis distances and angles, *Multivariate analysis IV* pp. 495–511.

Mardia, K., Kent, J. and Bibby, J. (1980). Multivariate analysis.

Mielke Jr, P. W. (1975). Convenient beta distribution likelihood techniques for describing and comparing meteorological data, *Journal of Applied Meteorology* **14**(6): 985–990.

Ord, J. K. (1972). *Families of frequency distributions*, Vol. 30, Griffin London.

Serdobolskii, V. (2000). *Multivariate statistical analysis: A high-dimensional approach*, Vol. 41, Springer.