Sri Lankan Journal of Applied Statistics (Special Issue)
Modern Statistical Methodologies in the Cutting Edge of Science

Asymptotic Distribution of the Largest Eigenvalue via Geometric Representations of High-Dimension, Low-Sample-Size Data

Aki Ishii[1*], Kazuyoshi Yata[2], and Makoto Aoshima[2]

[1]Graduate School of Pure and Applied Sciences, University of Tsukuba, Ibaraki, Japan

[2]Institute of Mathematics, University of Tsukuba, Ibaraki, Japan

[*]Corresponding Author: ishii-akitk@math.tsukuba.ac.jp

Received: 5, August 2013 / Revised: 28, January 2014 / Accepted: 18, February 2014

ABSTRACT

A common feature of high-dimensional data is that the data dimension is high, however, the sample size is relatively low. We call such a data HDLSS data. In this paper, we study HDLSS asymptotics for Gaussian-type HDLSS data. We find a surprising geometric representation of the HDLSS data in a dual space. We give an estimator of the eigenvalue by using the noise-reduction (NR) methodology. We show that the estimator enjoys consistency properties under mild conditions when the dimension is high. We provide an asymptotic distribution for the largest eigenvalue when the dimension is high while the sample size is fixed. We show that the estimator given by the NR methodology holds the asymptotic distribution under a condition milder than that for the conventional estimator.

Keywords: Dual space, HDLSS, Large p; small n, Noise-reduction methodology.

1. Introduction

A common feature of high-dimensional data is that the data dimension is high, however, the sample size is relatively low. This is the so-called "HDLSS" or "large p, small n" data situation where $p/n \to \infty$; here p is the data dimension and n is the sample size. The asymptotic studies of this type of data are becoming increasingly relevant. In recent years, substantial work had been done on the asymptotic behavior of eigenvalues of the sample covariance matrix in the limit as $p \to \infty$, see Johnstone (2001) and Paul (2007) for Gaussian data and Baik and Silverstein (2006) for non-Gaussian, i.i.d. data. Those literatures handled the cases when p and n increase at the same rate, i.e. $p/n \to c > 0$. The asymptotic behaviors of high-dimensional, low-sample-size (HDLSS) data were studied by Hall et al. (2005), Ahn et al. (2007), and Yata and Aoshima (2012) when $p \to \infty$ while n is fixed. They explored conditions to give a geometric representation of HDLSS data. The HDLSS asymptotic study usually assumes either the normality as the population distribution or a ρ-mixing condition as the dependency of random variables in a sphered data matrix. For instance, see Jung and Marron (2009). Yata and Aoshima (2009, 2013b) succeeded in investigating the consistency properties of both eigenvalues and eigenvectors in a more general framework. Yata and Aoshima (2012) gave consistent estimators of both the eigenvalues and eigenvectors together with the principal component (PC) scores by a method called the *noise-reduction (NR) methodology*. Yata and Aoshima (2010, 2013a) created the *cross-data-matrix (CDM) methodology* that provides a nonparametric method for non-Gaussian HDLSS data. On the other hand, Aoshima and Yata (2011a,b, 2013a) developed a variety of inference for high-dimensional data such as given-bandwidth confidence region, two-sample test, test of equality of two covariance matrices, classification, variable selection, regression, pathway analysis and so on along with sample size determination to ensure prespecified accuracy for each inference. See Aoshima and Yata (2013b,c) for a review covering this field of research.

In this paper, suppose we have a $p \times n$ data matrix, $\boldsymbol{X}_{(p)} = [\boldsymbol{x}_{1(p)}, ..., \boldsymbol{x}_{n(p)}]$, where $\boldsymbol{x}_{j(p)} = (x_{1j(p)}, ..., x_{pj(p)})^T,\ j = 1, ..., n$, are independent and identically distributed (i.i.d.) as a p-dimensional distribution with mean vector $\boldsymbol{\mu}_p$ and covariance matrix $\boldsymbol{\Sigma}_p\ (\geq \boldsymbol{O})$. We assume $n \geq 3$. The eigen-decomposition of $\boldsymbol{\Sigma}_p$ is given by $\boldsymbol{\Sigma}_p = \boldsymbol{H}_p\boldsymbol{\Lambda}_p\boldsymbol{H}_p^T$, where $\boldsymbol{\Lambda}_p$ is a diagonal matrix of eigenvalues, $\lambda_{1(p)} \geq \cdots \geq \lambda_{p(p)}(\geq 0)$, and $\boldsymbol{H}_p = [\boldsymbol{h}_{1(p)}, ..., \boldsymbol{h}_{p(p)}]$ is an orthogonal matrix of the corresponding eigenvectors. Let $\boldsymbol{X}_{(p)} - [\boldsymbol{\mu}_p, ..., \boldsymbol{\mu}_p] = \boldsymbol{H}_p\boldsymbol{\Lambda}_p^{1/2}\boldsymbol{Z}_{(p)}$. Then, $\boldsymbol{Z}_{(p)}$ is a $p \times n$ sphered data matrix from a distribution with the zero mean vector and the identity covariance matrix. Here, we write $\boldsymbol{Z}_{(p)} = [\boldsymbol{z}_{1(p)}, ..., \boldsymbol{z}_{p(p)}]^T$ and $\boldsymbol{z}_{j(p)} = (z_{j1(p)}, ..., z_{jn(p)})^T,\ j = 1, ..., p$. Note that $E(z_{ji(p)}z_{j'i(p)}) = 0\ (j \neq j')$ and $\mathrm{Var}(\boldsymbol{z}_{j(p)}) = \boldsymbol{I}_n$, where $\boldsymbol{I}_n$ is the n-dimensional identity matrix. Hereafter, the subscript p will be omitted for the sake of simplicity when it does not cause any confusion. We assume that the fourth moments of each variable in $\boldsymbol{Z}$ are uniformly bounded. Note that if $\boldsymbol{X}$ is Gaussian, z_{ij}s are i.i.d. as $N(0, 1)$, where $N(0, 1)$ denotes the standard normal distribution.

In this paper, we study HDLSS asymptotics for Gaussian-type HDLSS data when $p \to \infty$ while n is fixed. In Section 2, we find a surprising geometric representation of the HDLSS data in a dual space. In Section 3, we give an estimator of the eigenvalue by using the NR methodology. We show that the estimator enjoys consistency properties under mild conditions when the dimension is high. We provide an asymptotic distribution for the largest eigenvalue when the dimension is high while the sample size is fixed. We show that the estimator given by the NR methodology holds the asymptotic distribution under a condition milder than that for the conventional estimator. In Section 4, we summarize simulation studies of the findings.

2. Geometric Representations in a Dual Space

2.1 When $\boldsymbol{\mu}$ is Known

We assume $\boldsymbol{\mu} = \mathbf{0}$ without loss of generality. Let us write the sample covariance matrix as $\boldsymbol{S}_o = n^{-1}\boldsymbol{X}\boldsymbol{X}^T$. Then, we define the $n \times n$ dual sample covariance matrix by $\boldsymbol{S}_{oD} = n^{-1}\boldsymbol{X}^T\boldsymbol{X}$. Let $\hat{\lambda}_{o1} \geq \cdots \geq \hat{\lambda}_{on} \geq 0$ be the eigenvalues of $\boldsymbol{S}_{oD}$. Then, we define the eigen-decomposition of $\boldsymbol{S}_{oD}$ by $\boldsymbol{S}_{oD} = \sum_{j=1}^{n} \hat{\lambda}_{oj}\hat{\boldsymbol{u}}_{oj}\hat{\boldsymbol{u}}_{oj}^T$. Note that $\boldsymbol{S}_o$ and $\boldsymbol{S}_{oD}$ share non-zero eigenvalues. We consider the following condition.

$$\textbf{(A-i)} \qquad \frac{\text{tr}(\boldsymbol{\Sigma}^2)}{\text{tr}(\boldsymbol{\Sigma})^2} = \frac{\sum_{s=1}^{p} \lambda_s^2}{(\sum_{s=1}^{p} \lambda_s)^2} \to 0, \quad p \to \infty.$$

Note that (A-i) is equivalent to the condition that $\lambda_1/\text{tr}(\boldsymbol{\Sigma}) \to 0, p \to \infty$. Then, when $\boldsymbol{X}$ is Gaussian or $\boldsymbol{Z}$ is ρ-mixing, Ahn et al. (2007) and Jung and Marron (2009) showed a geometric representation as follows:

$$\frac{n}{\text{tr}(\boldsymbol{\Sigma})}\boldsymbol{S}_{oD} \xrightarrow{P} \boldsymbol{I}_n, \quad p \to \infty. \tag{2.1}$$

Let $\boldsymbol{w}_{oj} = \{n/\text{tr}(\boldsymbol{\Sigma})\}\hat{\lambda}_{oj}\hat{\boldsymbol{u}}_{oj}$ and $\boldsymbol{R}_{on} = \{\boldsymbol{e}_n \in \boldsymbol{R}^n | \, ||\boldsymbol{e}_n|| = 1\}$. Yata and Aoshima (2012) showed that

$$\boldsymbol{w}_{oj} \in \boldsymbol{R}_{on}, \quad j = 1, ..., n \tag{2.2}$$

in probability as $p \to \infty$. On the other hand, when $\boldsymbol{X}$ is non-Gaussian and $\boldsymbol{Z}$ is non-ρ-mixing, Yata and Aoshima (2012) showed another geometric representation as follows:

$$\frac{n}{\text{tr}(\boldsymbol{\Sigma})}\boldsymbol{S}_{oD} \xrightarrow{P} \boldsymbol{D}_n, \quad p \to \infty \tag{2.3}$$

where $\boldsymbol{D}_n$ is a diagonal matrix whose diagonal elements are of $O_P(1)$. Yata and Aoshima (2012) considered a boundary condition between (2.1) and (2.3) as follows:

$$\textbf{(A-ii)} \quad \frac{\text{Var}(||\boldsymbol{x}_k - \boldsymbol{\mu}||^2)}{\text{tr}(\boldsymbol{\Sigma})^2} = \frac{\sum_{r,s\ge 1}^{p} \lambda_r \lambda_s E\{(z_{rk}^2 - 1)(z_{sk}^2 - 1)\}}{(\sum_{s=1}^{p} \lambda_s)^2} \to 0, \quad p \to \infty.$$

Then, they gave the following result.

Theorem 2.1 (Yata and Aoshima, 2012)**.** *Assume (A-i). If the elements of $\boldsymbol{Z}$ satisfy (A-ii), we have (2.1) as $p \to \infty$. Otherwise, we have (2.3) as $p \to \infty$.*

2.2 When $\boldsymbol{\mu}$ is Unknown

Let us write the sample covariance matrix as $\boldsymbol{S} = (n-1)^{-1}(\boldsymbol{X} - \overline{\boldsymbol{X}})(\boldsymbol{X} - \overline{\boldsymbol{X}})^T = (n-1)^{-1}\sum_{j=1}^{n}(\boldsymbol{x}_j - \bar{\boldsymbol{x}})(\boldsymbol{x}_j - \bar{\boldsymbol{x}})^T$, where $\overline{\boldsymbol{X}} = [\bar{\boldsymbol{x}}, ..., \bar{\boldsymbol{x}}]$ and $\bar{\boldsymbol{x}} = \sum_{j=1}^{n} \boldsymbol{x}_j/n$. Then, we define the $n \times n$ dual sample covariance matrix by $\boldsymbol{S}_D = (n-1)^{-1}(\boldsymbol{X} - \overline{\boldsymbol{X}})^T(\boldsymbol{X} - \overline{\boldsymbol{X}})$. Let $\hat{\lambda}_1 \ge \cdots \ge \hat{\lambda}_{n-1} \ge 0$ be the eigenvalues of $\boldsymbol{S}_D$. Let us write the eigen-decomposition of $\boldsymbol{S}_D$ as $\boldsymbol{S}_D = \sum_{j=1}^{n-1} \hat{\lambda}_j \hat{\boldsymbol{u}}_j \hat{\boldsymbol{u}}_j^T$. Note that $\boldsymbol{S}$ and $\boldsymbol{S}_D$ share non-zero eigenvalues. Then, we have the following geometric representation for $\boldsymbol{S}_D$.

Theorem 2.2. *Assume (A-i) and (A-ii). Then, we have as $p \to \infty$ that*

$$\frac{n-1}{tr(\boldsymbol{\Sigma})}\boldsymbol{S}_D \xrightarrow{P} \boldsymbol{I}_n - \frac{1}{n}\boldsymbol{1}_n\boldsymbol{1}_n^T,$$

where $\boldsymbol{1}_n = (1, ..., 1)^T$.

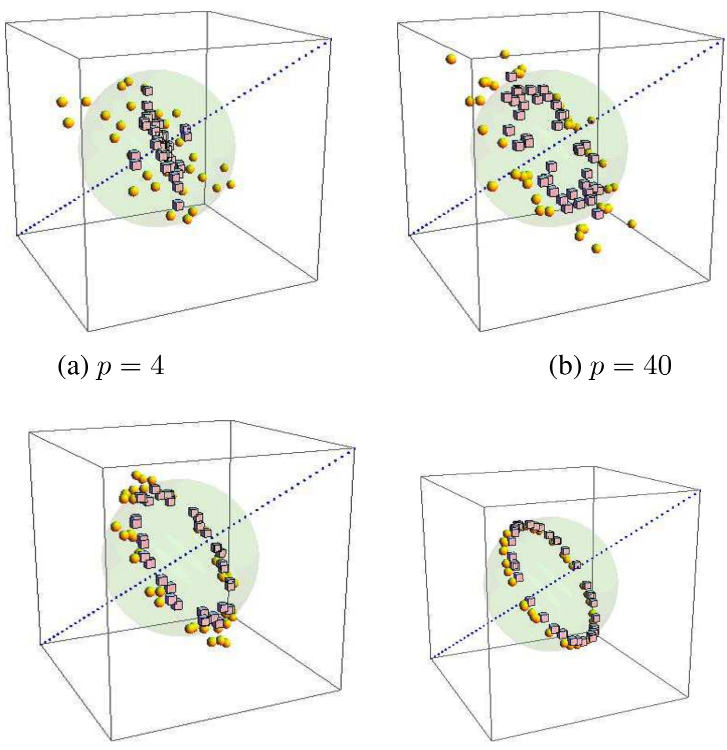
Corollary 2.1. *Let $\boldsymbol{w}_j = \{(n-1)/tr(\boldsymbol{\Sigma})\}\hat{\lambda}_j\hat{\boldsymbol{u}}_j$. Assume (A-i) and (A-ii). Then, we have as $p \to \infty$ that*

$$\frac{(n-1)\hat{\lambda}_j}{tr(\boldsymbol{\Sigma})} = \frac{(n-1)\hat{\boldsymbol{u}}_j^T\boldsymbol{S}_D\hat{\boldsymbol{u}}_j}{tr(\boldsymbol{\Sigma})} \xrightarrow{P} 1, \quad j = 1, ..., n-1;$$
$$\boldsymbol{w}_j \in \boldsymbol{R}_n, \quad j = 1, ..., n-1$$

in probability as $p \to \infty$, where $\boldsymbol{R}_n = \{\boldsymbol{e}_n \in \boldsymbol{R}^n |\ \boldsymbol{e}_n^T\boldsymbol{1}_n = 0,\ ||\boldsymbol{e}_n|| = 1\}$.

From Corollary 2.1 the eigenspace spanned by $\hat{\boldsymbol{u}}_j,\ j = 1, ..., n-1$, is close to the orthogonal complement of $\boldsymbol{1}_n$ in $\boldsymbol{R}^n$ as $p \to \infty$ and the direction of the eigenvectors is not uniquely determined. On the other hand, the eigenvalues become deterministic but there becomes no difference among them. For these reasons, it is difficult to estimate the eigenvalues and the eigenvectors by using $\boldsymbol{S}_D$ (or $\boldsymbol{S}$) in conventional PCA.

Let us observe a geometric representation given by Corollary 2.1. Now, we consider an easy example such as $\lambda_1 = \cdots = \lambda_p = 1$ and $n = 3$. In Fig. 2.1, we displayed scatter plots of 20 independent pairs of $\pm \boldsymbol{w}_j$ $(j = 1, 2)$ that were generated from $N_p(\boldsymbol{\mu}, \boldsymbol{I}_p)$ for (a) $p = 4$, (b) $p = 40$, (c) $p = 400$ and (d) $p = 4000$. We denoted $\boldsymbol{w}_1$ by and $\boldsymbol{w}_2$ by . We also denoted $\mathbf{1}_n = (1, 1, 1)^T$ by the dotted line. We observed that all the plots of $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ gather on the surface of the orthogonal complement of $\mathbf{1}_n = (1, 1, 1)^T$ in $\boldsymbol{R}^3$ when p is large. Moreover, they appeared around the unit circle on the orthogonal complement of $\mathbf{1}_n = (1, 1, 1)^T$ in $\boldsymbol{R}^3$ as expected by Corollary 2.1



(a) $p = 4$ (b) $p = 40$

(c) $p = 400$ (d) $p = 4000$

Figure 2.1: The geometric representation of 20 pairs of $\pm \boldsymbol{w}_j (j = 1, 2)$ from $N_p(\boldsymbol{\mu}, \boldsymbol{I}_p)$ when $p = 4, 40, 400$ and 4000. We denoted $\boldsymbol{w}_1$ by , $\boldsymbol{w}_2$ by and $\mathbf{1}_n = (1, 1, 1)^T$ by the dotted line.

3. Largest Eigenvalue and Its Asymptotic Distribution

In this section, we consider eigenvalue estimation and give an asymptotic distribution for the largest eigenvalue. Yata and Aoshima (2012) proposed a method for eigenvalue estimation called the *noise-reduction (NR) methodology* that was brought by the geometric representation in (2.2). When we apply the NR methodology to the case when $\boldsymbol{\mu}$ is unknown, the NR estimator of λ_i is given by

$$\tilde{\lambda}_i = \hat{\lambda}_i - \frac{\text{tr}(\boldsymbol{S}_D) - \sum_{j=1}^{i} \hat{\lambda}_j}{n-1-i} \quad (i = 1, ..., n-2). \tag{3.1}$$

Note that $\tilde{\lambda}_i \geq 0$ for $i = 1, ..., n-2$. Yata and Aoshima (2012, 2013b) showed that $\tilde{\lambda}_j$ has several consistency properties when $p \to \infty$ and $n \to \infty$. In this paper, we focus on the largest eigenvalue, $\tilde{\lambda}_1$, that has the most important information in data analysis and study its asymptotic properties when $p \to \infty$ while n is fixed. We assume the following conditions for the population largest eigenvalue:

(A-iii) $$\frac{\text{tr}(\boldsymbol{\Sigma}^2) - \lambda_1^2}{\lambda_1^2} = \frac{\sum_{s=2}^{p} \lambda_s^2}{\lambda_1^2} \to 0, \quad p \to \infty;$$

(A-iv) $$\frac{\sum_{r,s \geq 2}^{p} \lambda_r \lambda_s E\{(z_{rk}^2 - 1)(z_{sk}^2 - 1)\}}{\lambda_1^2} \to 0, \quad p \to \infty.$$

Note that (A-iv) is naturally satisfied for the case when X is Gaussian and (A-iii) is met. Let $\boldsymbol{z}_{oj} = \boldsymbol{z}_j - (\bar{z}_j, ..., \bar{z}_j)^T$, $j = 1, ..., p$, where $\bar{z}_j = n^{-1} \sum_{k=1}^{n} z_{jk}$. We write that

$$(n-1)\boldsymbol{S}_D = \lambda_1 \boldsymbol{z}_{o1} \boldsymbol{z}_{o1}^T + \sum_{j=2}^{p} \lambda_j \boldsymbol{z}_{oj} \boldsymbol{z}_{oj}^T.$$

Then, from Corollary 2.1, under (A-iii) and (A-iv), we have $p \to \infty$ that

$$\hat{\boldsymbol{u}}_i^T \frac{\sum_{j=2}^{p} \lambda_j \boldsymbol{z}_{oj} \boldsymbol{z}_{oj}^T}{\lambda_1} \hat{\boldsymbol{u}}_i = \frac{\text{tr}(\boldsymbol{\Sigma}) - \lambda_1}{\lambda_1} + o_p(1), \quad i = 1, ..., n-1.$$

Therefore, we have that

$$\frac{\hat{\lambda}_1}{\lambda_1} = \hat{\boldsymbol{u}}_1^T \frac{\boldsymbol{S}_D}{\lambda_1} \hat{\boldsymbol{u}}_1 = (\hat{\boldsymbol{u}}_1^T \boldsymbol{z}_{o1}/\sqrt{n-1})^2 + \frac{\text{tr}(\boldsymbol{\Sigma}) - \lambda_1}{\lambda_1(n-1)} + o_p(1).$$

Note that $\boldsymbol{z}_{o1}^T \boldsymbol{1}_n = 0$. If $P(\lim_{p \to \infty} ||\boldsymbol{z}_{o1}|| \neq 0) = 1$, we have as $p \to \infty$ that

$$\frac{\hat{\lambda}_1}{\lambda_1} = \hat{\boldsymbol{u}}_1^T \frac{\boldsymbol{S}_D}{\lambda_1} \hat{\boldsymbol{u}}_1 = ||\boldsymbol{z}_{o1}/\sqrt{n-1}||^2 + \frac{\text{tr}(\boldsymbol{\Sigma}) - \lambda_1}{\lambda_1(n-1)} + o_p(1) \tag{3.2}$$

under (A-iii) and (A-iv).

For the second term in (3.1) with $i = 1$, we have the following result.

Lemma 3.1. *Assume* $P(\lim_{p\to\infty} ||\boldsymbol{z}_{o1}|| \neq 0) = 1$. *Under (A-iii) and (A-iv), it holds as* $p \to \infty$ *that*

$$\frac{tr(\boldsymbol{\Sigma}) - \lambda_1}{\lambda_1(n-1)} - \frac{tr(\boldsymbol{S}_D) - \hat{\lambda}_1}{\lambda_1(n-2)} = o_p(1).$$

Then, from (3.2) and Lemma 3.1, we have the following results.

Theorem 3.1. *Assume* $P(\lim_{p\to\infty} ||\boldsymbol{z}_{o1}|| \neq 0) = 1$. *Under (A-iii) and (A-iv), it holds as* $p \to \infty$ *that*

$$\frac{\tilde{\lambda}_1}{\lambda_1} = ||\boldsymbol{z}_{o1}/\sqrt{n-1}||^2 + o_p(1).$$

Corollary 3.1. *If* $z_{1j},\ j = 1, .., n$, *are i.i.d. as* $N(0, 1)$, *it holds as* $p \to \infty$ *that*

$$(n-1)\frac{\tilde{\lambda}_1}{\lambda_1} \Rightarrow \chi^2_{n-1}$$

under (A-iii) and (A-iv). Here, " $\Rightarrow$ *" denotes the convergence in distribution and* χ^2_{n-1} *denotes a random variable distributed as* χ^2 *distribution with* $n-1$ *degrees of freedom.*

Next, we consider asymptotic properties of the conventional estimator, $\hat{\lambda}_1$, for the sake of comparison when $p \to \infty$ while n is fixed. We assume the following condition for the population largest eigenvalue:

$$\textbf{(A-v)} \qquad \frac{\text{tr}(\boldsymbol{\Sigma}) - \lambda_1}{\lambda_1} = \frac{\sum_{s=2}^{p} \lambda_s}{\lambda_1} \to 0, \quad p \to \infty.$$

Under (A-v), it holds that $\sum_{s=2}^{p} \lambda_s^2/\lambda_1^2 \leq \lambda_2 \sum_{s=2}^{p} \lambda_s/\lambda_1^2 \leq \sum_{s=2}^{p} \lambda_s/\lambda_1 \to 0$ and $\sum_{r,s\geq 2}^{p} \lambda_r\lambda_s E\{(z_{rk}^2 - 1)(z_{sk}^2 - 1)\}/\lambda_1^2 = O\{(\sum_{s=2}^{p} \lambda_s)^2/\lambda_1^2\} \to 0$. Hence, (A-v) is stronger than the conditions (A-iii) and (A-iv). From (3.2), for the conventional estimator $\hat{\lambda}_1$, we have the following result.

Corollary 3.2. *Assume* $P(\lim_{p\to\infty} ||\boldsymbol{z}_{o1}|| \neq 0) = 1$. *Under (A-v), it holds as* $p \to \infty$ *that*

$$\frac{\hat{\lambda}_1}{\lambda_1} = ||\boldsymbol{z}_{o1}/\sqrt{n-1}||^2 + o_p(1).$$

In addition, if $z_{1j}, j = 1, ..., n$, *are i.i.d. as* $N(0, 1)$, *it holds that*

$$(n-1)\frac{\hat{\lambda}_1}{\lambda_1} \Rightarrow \chi^2_{n-1}. \tag{3.3}$$

Remark 3.1. Jung and Marron (2009) gave (3.3) under different but still strict assumptions.

By comparing Theorem 3.1 and Corollary 3.1 with Corollary 3.2, we can conclude that $\tilde{\lambda}_1$ has the asymptotic properties under milder conditions than $\hat{\lambda}_1$ when $p \to \infty$ while n is fixed. In fact, (A-v) is a too strict condition in real high-dimensional data analyses. It should be noted that (A-v) is equivalent to the condition that $\lambda_1/\text{tr}(\boldsymbol{\Sigma}) \to 1,\ p \to \infty$, that is (A-v) means that the contribution ratio of the first principal component is asymptotically 1 as $p \to \infty$.

Let $\text{Var}(z_{1k}^2) = M_1\ (< \infty)$ and assume $\liminf_{p\to\infty} M_1 > 0$. Note that $M_1 = 2$ if $z_{1j},\ j = 1, ..., n$, are i.i.d. as $N(0, 1)$. When $p \to \infty$ and $n \to \infty$, we have the following result from Yata and Aoshima (2013b).

Theorem 3.2 (Yata and Aoshima, 2013b)**.** *Under (A-iii) and (A-iv), it holds as* $p \to \infty$ *and* $n \to \infty$ *that*

$$\sqrt{\frac{n-1}{M_1}}\Big(\frac{\tilde{\lambda}_1}{\lambda_1} - 1\Big) \Rightarrow N(0, 1).$$

Remark 3.2. Under (A-v), it holds as $p \to \infty$ and $n \to \infty$ that

$$\sqrt{\frac{n-1}{M_1}}\Big(\frac{\hat{\lambda}_1}{\lambda_1} - 1\Big) \Rightarrow N(0, 1).$$

4. SIMULATION

In order to study the distributions of $\tilde{\lambda}_1$ and $\hat{\lambda}_1$, we used computer simulations. We set $\boldsymbol{\Sigma} = \text{diag}(\lambda_1, ..., \lambda_p)$ with $\lambda_1 = p^{2/3}$ and $\lambda_2 = \cdots = \lambda_p = 1$. We considered the cases of $p =$ 20, 100, 500 and 2500 when (a) $n = 5$ and (b) $n = 25$. We generated $\boldsymbol{x}_j$, $j = 1, ..., n$, independently from the p-dimensional normal distribution, $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Note that (A-iii) and (A-iv) hold, however, (A-v) does not hold. We denoted independent pseudorandom 2000 observations of $\tilde{\lambda}_1$ and $\hat{\lambda}_1$ by $\tilde{\lambda}_{1r}$ and $\hat{\lambda}_{1r}$ for $r = 1, ..., 2000$. In the end of the rth replication, we checked whether the event, $(n-1)\tilde{\lambda}_{1r}/\lambda_1 \le a_{n-1}$, is true (or false) and defined $P_{ir} = 1$ (or 0) accordingly, where a_{n-1} is the upper 0.05 point of χ^2_{n-1}. We calculated $\overline{P}(0.95) = \sum_{r=1}^{2000} P_r/2000$ as an estimate of $P\{(n-1)\tilde{\lambda}_1/\lambda_1 \le a_{n-1}\}$. Note that the standard deviation of the estimates is less than 0.011. As for $\hat{\lambda}_1$ as well, we calculated $\overline{P}(0.95) = \sum_{r=1}^{2000} P_r/2000$ similarly as an estimate of $P\{(n-1)\hat{\lambda}_1/\lambda_1 \le a_{n-1}\}$.

In Fig. 4.1, we gave the histograms of $(n-1)\tilde{\lambda}_1/\lambda_1$ (left panel) and $(n-1)\hat{\lambda}_1/\lambda_1$ (right panel) together with $\overline{P}(0.95)$ for $p =$ 20, 100, 500 and 2500 when (a) $n = 5$ and (b) $n = 25$. From Corollaries 3.1 and 3.2, we displayed the asymptotic probability density of $(n-1)\tilde{\lambda}_1/\lambda_1$ (or $(n-1)\hat{\lambda}_1/\lambda_1$), χ^2_{n-1}. We observed that the histograms of $(n-1)\tilde{\lambda}_1/\lambda_1$ become close to χ^2_{n-1} as p increases even when $n = 5$. On the other hand, the histograms of $(n-1)\hat{\lambda}_1/\lambda_1$ became separated from χ^2_{n-1} as p increases especially when $n = 5$. That is because the second term in (3.2) becomes large as p increases. The NR estimator, $\tilde{\lambda}_1$, gives a good approximation to the asymptotic distribution in such a case as well by removing the term as in (3.1).

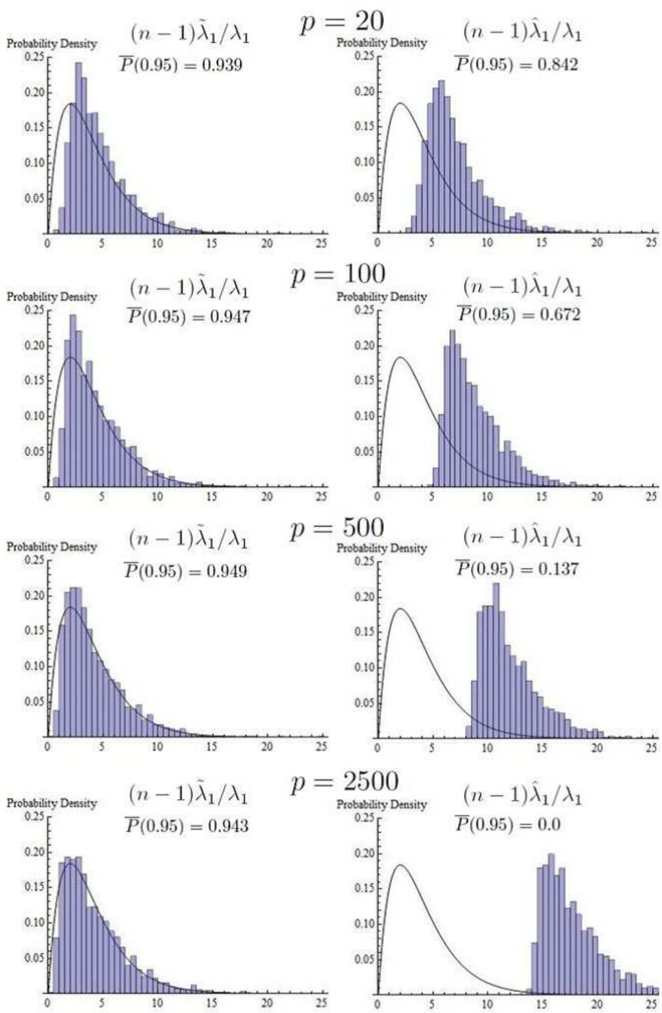


Figure 4.1: (a) When $n = 5$. The histograms of $(n-1)\tilde{\lambda}_1/\lambda_1$ (left panels) and $(n-1)\hat{\lambda}_1/\lambda_1$ (right panels) together with the probability density of χ^2_{n-1} and $\overline{P}(0.95)$ for $p = 20, 100, 500$ and 2500.

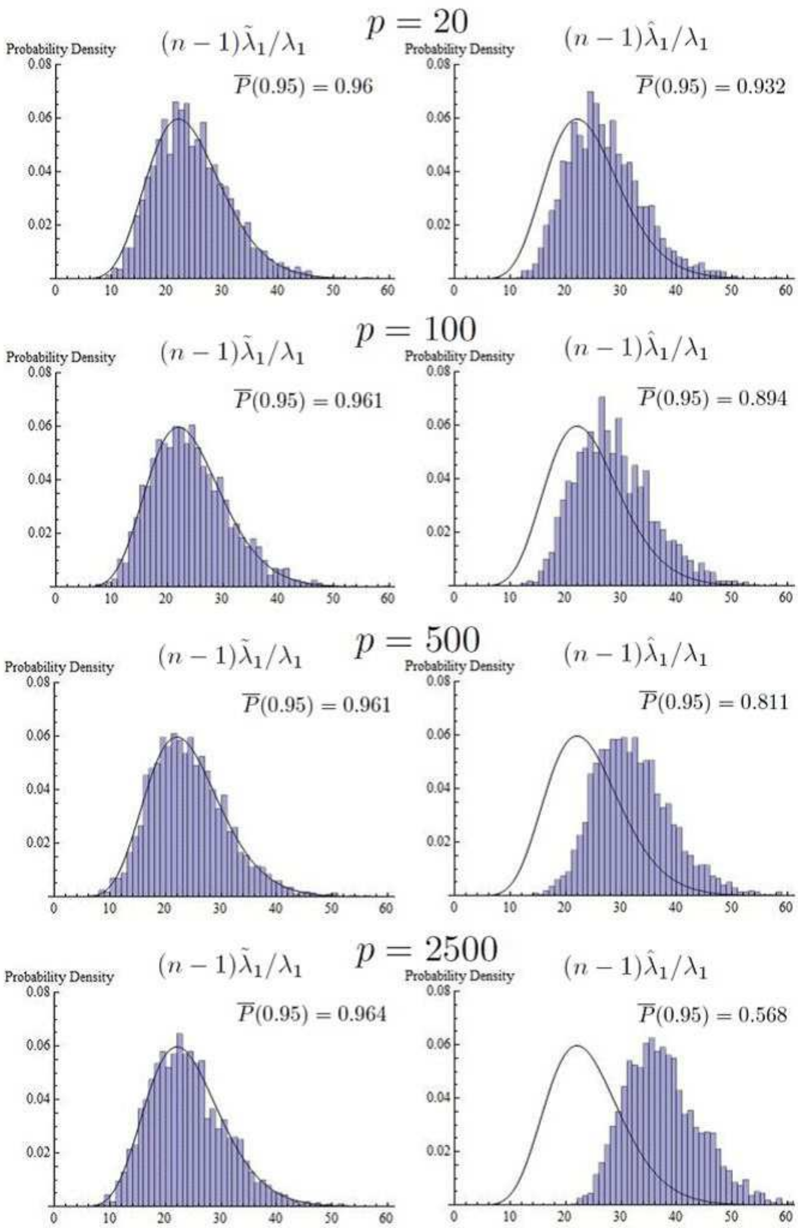


Figure 4.1 (Continued): (b) When $n = 25$.

Appendix

Proof of Theorem 2.2. By using Chebyshev's inequality, for any $\tau > 0$, we have as $p \to \infty$ that

$$P\Big(\Big|\frac{||\boldsymbol{x}_k - \boldsymbol{\mu}||^2}{\text{tr}(\boldsymbol{\Sigma})} - 1\Big| > \tau\Big) \le \tau^{-2}\frac{\sum_{r,s\ge1}^{p}\lambda_r\lambda_s E\{(z_{rk}^2-1)(z_{sk}^2-1)\}}{\text{tr}(\boldsymbol{\Sigma})^2} \to 0;$$
$$P\Big(\Big|\frac{(\boldsymbol{x}_k - \boldsymbol{\mu})^T(\boldsymbol{x}_{k'} - \boldsymbol{\mu})}{\text{tr}(\boldsymbol{\Sigma})}\Big| > \tau\Big) \le \tau^{-2}\frac{\text{tr}(\boldsymbol{\Sigma}^2)}{\text{tr}(\boldsymbol{\Sigma})^2} \to 0 \quad (k \ne k')$$

under (A-i) and (A-ii). Then, we have $(\boldsymbol{X}-[\boldsymbol{\mu},...,\boldsymbol{\mu}])^T(\boldsymbol{X}-[\boldsymbol{\mu},...,\boldsymbol{\mu}])/\text{tr}(\boldsymbol{\Sigma}) \xrightarrow{P} \boldsymbol{I}_n$. Note that $(\boldsymbol{X} - [\boldsymbol{\mu},...,\boldsymbol{\mu}])(\boldsymbol{I}_n - \boldsymbol{1}_n\boldsymbol{1}_n^T/n) = \boldsymbol{X} - \overline{\boldsymbol{X}}$. Thus we write that

$$\boldsymbol{S}_D = \frac{(\boldsymbol{I}_n - \boldsymbol{1}_n\boldsymbol{1}_n^T/n)(\boldsymbol{X} - [\boldsymbol{\mu},...,\boldsymbol{\mu}])^T(\boldsymbol{X} - [\boldsymbol{\mu},...,\boldsymbol{\mu}])(\boldsymbol{I}_n - \boldsymbol{1}_n\boldsymbol{1}_n^T/n)}{n-1}.$$

Hence, we have that

$$\frac{(n-1)\boldsymbol{S}_D}{\text{tr}(\boldsymbol{\Sigma})} \xrightarrow{P} \boldsymbol{I}_n - \frac{1}{n}\boldsymbol{1}_n\boldsymbol{1}_n^T.$$

It concludes the result. □

Proof of Corollary 2.1. From Theorem 2.2, it follows that $\text{rank}(\boldsymbol{S}_D) = n-1$ asymptotically. By noting that $\hat{\boldsymbol{u}}_i^T\boldsymbol{1}_n = 0$ with probability tending to 1 for $i = 1,...,n-1$, it concludes the results. □

Proof of Lemma 3.1. By using Chebyshev's inequality, under (A-iii) and (A-iv), we have as $p \to \infty$ that

$$\begin{aligned}
&P\Big(\Big|\frac{\text{tr}(\boldsymbol{S}_D)}{\lambda_1} - \frac{\text{tr}(\boldsymbol{\Sigma}) - \lambda_1 + \lambda_1||\boldsymbol{z}_{o1}/\sqrt{n-1}||^2}{\lambda_1}\Big| > \tau\Big)\\
&= P\Big(\Big|\frac{\sum_{s=2}^{p}\lambda_s\{||\boldsymbol{z}_{os}||^2 - (n-1)\}}{\lambda_1(n-1)}\Big| > \tau\Big)\\
&= O\Big(\frac{\sum_{r\ne s\ge2}^{p}\lambda_r\lambda_s E\{(z_{rk}^2-1)(z_{sk}^2-1)\}}{\lambda_1^2}\Big) + O\Big(\frac{\sum_{s=2}^{p}\lambda_s^2}{\lambda_1^2}\Big) \to 0,
\end{aligned}$$

so that $\text{tr}(\boldsymbol{S}_D)/\lambda_1 = (\text{tr}(\boldsymbol{\Sigma}) - \lambda_1 + \lambda_1||\boldsymbol{z}_{o1}/\sqrt{n-1}||^2)/\lambda_1 + o_p(1)$. Therefore, from (3.2), we can claim that

$$\begin{aligned}
\frac{\text{tr}(\boldsymbol{S}_D) - \hat{\lambda}_1}{\lambda_1(n-2)} &= \frac{\text{tr}(\boldsymbol{\Sigma}) - \lambda_1 + \lambda_1||\boldsymbol{z}_{o1}/\sqrt{n-1}||^2 - \hat{\lambda}_1}{\lambda_1(n-2)} + o_p(1)\\
&= \frac{\text{tr}(\boldsymbol{\Sigma}) - \lambda_1}{\lambda_1(n-1)} + o_p(1)
\end{aligned}$$

under (A-iii), (A-iv) and $P(\lim_{p\to\infty} ||\boldsymbol{z}_{o1}|| \neq 0) = 1$. It concludes result. □

Proofs of Theorem 3.1 and Corollary 3.1. By combining Lemma 3.1 with (3.2), we can claim the result of Theorem 3.1 straightforwardly. Next, we consider Corollary 3.1. We note that $||\boldsymbol{z}_{o1}||^2 = \sum_{k=1}^{n} z_{1k}^2 - n\bar{z}_1^2$ is distributed as χ_{n-1}^2 if $z_{1j},\ j = 1, ..., n$, are i.i.d. as $N(0, 1)$. From the fact that $P(\chi_{n-1}^2 \neq 0) = 1$, it concludes the results. □

Proof of Corollary 3.2. Note that (A-v) implies (A-iii) and (A-iv). Thus in a way similar to the proofs of Theorem 3.1 and Corollary 3.1, we can conclude the results.

Acknowledgements

Research of the second author was partially supported by Grant-in-Aid for Young Scientists (B), Japan Society for the Promotion of Science, JSPS, under Contract Number 23740066. Research of the third author was partially supported by Grants-in-Aid for Scientific Research (B) and Challenging Exploratory Research, JSPS, under Contract Numbers 22300094 and 23650142.

References

1. Ahn, J., Marron, J. S., Muller, K. M., and Chi, Y.-Y. (2007). The High-Dimension, Low-Sample-Size Geometric Representation Holds under Mild Conditions, *Biometrika* 94: 760-766. DOI: 10.1093/biomet/asm050

2. Aoshima, M. and Yata, K. (2011a). Two-Stage Procedures for High-Dimensional Data, *Sequential Analysis* (*Editor's special invited paper*) 30: 356-399. DOI: 10.1080/07474946.2011.619088

3. Aoshima, M. and Yata, K. (2011b). Authors' Response, *Sequential Analysis* 30: 432-440. DOI: 10.1080/07474946.2011.619102

4. Aoshima, M. and Yata, K. (2013a). Asymptotic Normality for Inference on Multisample, High-Dimensional Mean Vectors under Mild Conditions, *Methodology and Computing in Applied Probability*, in press. DOI: 10.1007/s11009-013-9370-7

5. Aoshima, M. and Yata, K. (2013b). Invited Review Article: Statistical Inference in High-Dimension, Low-Sample-Size Settings, *Sugaku* 65: 225-247.

6. Aoshima, M. and Yata, K. (2013c). The JSS Research Prize Lecture: Ef-

fective Methodologies for High-Dimensional Data, *Journal of the Japan Statistical Society, Series J* 43: 123-150.

7. Baik, J. and Silverstein. (2006). Eigenvalues of Large Sample Covariance Matrices of Spiked Population Models, *Journal of Multivariate Analysis* 97: 1382-1408. DOI: 10.1016/j.jmva.2005.08.003
8. Hall, P., Marron, J.S., and Neeman, A. (2005). Geometric Representation of High Dimension, Low Sample Size Data, *Journal of the Royal Statistical Society, Series B* 67: 427-444. DOI: 10.1111/j.1467-9868.2005.00510.x
9. Johnstone, I.M. (2001). On the Distribution of the Largest Eigenvalue in Principal Components Analysis, *The Annals of Statistics* 29: 295-327. DOI: 10.1214/aos/1009210544
10. Jung, S. and Marron, J.S. (2009). PCA Consistency in High Dimension, Low Sample Size Context, *The Annals of Statistics* 37: 4104-4130. DOI: 10.1214/09-AOS709
11. Paul, D. (2007). Asymptotics of Sample Eigenstructure for a Large Dimensional Spiked Covariance Model, *Statistica Sinica* 17: 1617-1642.
12. Yata, K. and Aoshima, M. (2009). PCA Consistency for Non-Gaussian Data in High Dimension, Low Sample Size Context, *Communications in Statistics -Theory & Methods*, Special Issue Honoring Zacks, S. (ed. Mukhopadhyay, N.) 38: 2634-2652. DOI: 10.1080/03610910902936083
13. Yata, K. and Aoshima, M. (2010). Effective PCA for High-Dimension, Low-Sample-Size Data with Singular Value Decomposition of Cross Data Matrix, *Journal of Multivariate Analysis* 101: 2060-2077. DOI: 10.1016/j.jmva.2010.04.006
14. Yata, K. and Aoshima, M. (2012). Effective PCA for High-Dimension, Low- Sample-Size Data with Noise Reduction via Geometric Representations, *Journal of Multivariate Analysis* 105: 193-215. DOI: 10.1016/j.jmva.2011.09.002
15. Yata, K. and Aoshima, M. (2013a). Correlation Tests for High-Dimensional Data using Extended Cross-Data-Matrix Methodology, *Journal of Multivariate Analysis* 117: 313-331. DOI: 10.1016/j.jmva.2013.03.007
16. Yata, K. and Aoshima, M. (2013b). PCA Consistency for the Power Spiked Model in High-Dimensional Settings, *Journal of Multivariate Analysis* 122: 334-354. DOI: 10.1016/j.jmva.2013.08.003