数理解析研究所講究録
第1910巻 2014年 115-124

Largest Eigenvalue Estimation for High-Dimension, Low-Sample-Size Data and its Application

Aki Ishii[1], Kazuyoshi Yata[2], 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

Abstract: A common feature of high-dimensional data is the data dimension is high, however, the sample size is relatively low. We call such data HDLSS data. In this paper, we study HDLSS asymptotics when the data dimension is high while the sample size is fixed. We first introduce two eigenvalue estimation methods: the noise-reduction (NR) methodology and the cross-data-matrix (CDM) methodology. We show that the eigenvalue estimators by the NR and the CDM enjoy asymptotic properties under mild conditions when the data dimension is high. We provide asymptotic distributions of those estimators in the HDLSS context where the data dimension is high while the sample size is fixed. We give a bias corrected CDM estimator of the largest eigenvalue. We show that the NR estimator has the asymptotic distribution under a mild condition and so does the bias corrected CDM estimator under a more relaxed condition. We give an application to construct confidence intervals of the first contribution ratio in the HDLSS context. Finally, we summarize simulation results.

Keywords: Contribution ratio; Cross-data-matrix methodology; HDLSS; Large p, small n; Noise-reduction methodology; Principal component analysis.

1 Introduction

One of the features of modern data is the data has a high dimension and a low sample size. We call such data "HDLSS" or "large p, small n" data where $p/n \to \infty$; here p is the data dimension and n is the sample size. The asymptotic behaviors of 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 several types of geometric representations 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. See Jung and Marron (2009). In a more general framework, Yata and Aoshima (2009) succeeded in proving consistency properties for both eigenvalues and eigenvectors of the sample covariance matrix and showed that the conventional principal component analysis (PCA) cannot give a consistent estimate in the HDLSS context. In order to overcome this inconvenience, Yata and Aoshima (2012) developed the *noise-reduction (NR) methodology* to give consistent estimators of both eigenvalues and eigenvectors together with principal component scores for Gaussian-type HDLSS data. As for non-Gaussian HDLSS data, Yata and Aoshima (2010, 2013) created the *cross-data-matrix (CDM) methodology* that provides a nonparametric method to ensure the consistent properties in the HDLSS context. On the other hand, Aoshima and Yata (2011a,b, 2013a) developed a variety of inference for HDLSS 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 $\boldsymbol{\Sigma}_p$ (≥ 0). We assume $n \geq 4$. 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 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. 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 $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_{n-1} \geq 0$ 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.

In this paper, we study HDLSS asymptotics when $p \to \infty$ while n is fixed. In Section 2, we show that the eigenvalue estimators by the NR and the CDM enjoy asymptotic properties under mild conditions when the data dimension is high. We provide asymptotic distributions of those estimators in the HDLSS context. We give a bias corrected CDM estimator of the largest eigenvalue. We show that the NR estimator has the asymptotic properties under a mild condition and so does the bias corrected CDM estimator under a more relaxed condition. In Section 3, we give an application to construct confidence intervals of the first contribution ratio in the HDLSS context. Finally, in Section 4, we summarize simulation results.

2 Largest Eigenvalue Estimation and its Asymptotic Distribution

In this section, we consider eigenvalue estimation and give an asymptotic distribution for the largest eigenvalue when $p \to \infty$ while n is fixed.

2.1 Noise-Reduction Estimator

Yata and Aoshima (2012) proposed a method for eigenvalue estimation called the *noise-reduction (NR) methodology* that was brought by a geometric representation. See Sections 2 and 3 in Yata and Aoshima (2012) for the details. When we apply the NR methodology, the NR estimator of λ_j is given by

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

Note that $\tilde{\lambda}_j \geq 0$ for $j = 1, ..., n-2$. Yata and Aoshima (2012, 2013) 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 analyses. We assume the following conditions for the largest eigenvalue:

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

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

Note that (A-ii) is naturally satisfied when X is Gaussian and (A-i) is met. Let $\boldsymbol{z}_{oj}=\boldsymbol{z}_j-(\bar{z}_j,...,\bar{z}_j)^T$, $j=1,...,p$, where $\bar{z}_j=n^{-1}\sum_{k=1}^{n}z_{jk}$. Then, Ishii et al. (2014) gave the following results.

Theorem 2.1 (Ishii et al., 2014). *Assume* $P(\lim_{p\to\infty}||\boldsymbol{z}_{o1}||\neq 0)=1$. *Under (A-i) and (A-ii), it holds that as* $p\to\infty$

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

Corollary 2.1 (Ishii et al., 2014). *If* z_{1j}, $j=1,...,n$, *are i.i.d. as* $N(0,1)$, *it holds that as* $p\to\infty$

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

under (A-i) and (A-ii). Here, " $\Rightarrow$ " denotes the convergence in distribution and χ_{n-1}^2 *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 largest eigenvalue:

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

Under (A-iii), 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-iii) is stronger than (A-i) and (A-ii). For the conventional estimator $\hat{\lambda}_1$, Ishii et al. (2014) gave the following results.

Corollary 2.2 (Ishii et al., 2014). *Assume* $P(\lim_{p\to\infty}||\boldsymbol{z}_{o1}||\neq 0)=1$. *Under (A-iii), it holds as* $p\to\infty$

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

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

$$(n-1)\frac{\hat{\lambda}_1}{\lambda_1}\Rightarrow\chi_{n-1}^2. \tag{2.1}$$

Remark 2.1. Jung and Marron (2009) gave (2.1) under different but still strict assumptions.

Remark 2.2. By comparing Theorem 2.1 and Corollary 2.1 with Corollary 2.2, 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-iii) is a too strict condition in real high-dimensional data analyses. It should be noted that (A-iii) is equivalent to the condition that $\lambda_1/\mathrm{tr}(\boldsymbol{\Sigma})\to 1$, $p\to\infty$, that is (A-iii) means that the contribution ratio of the first principal component is asymptotically 1 as $p\to\infty$.

2.2 Bias Corrected Cross-Data-Matrix Estimator

We consider the case when (A-ii) is not always met. In such cases, the NR methodology does not ensure the asymptotic properties. Yata and Aoshima (2010) proposed a method called the *cross-data-matrix (CDM) methodology* to proceed with eigenvalue estimation even in such cases. Let $n_1 = \lceil n/2 \rceil$ and $n_2 = n - n_1$, where $\lceil x \rceil$ denotes the smallest integer $\geq x$. We divide the data matrix $\boldsymbol{X}$ into $\boldsymbol{X}_1 = [\boldsymbol{x}_{11}, ..., \boldsymbol{x}_{1n_1}]$ and $\boldsymbol{X}_2 = [\boldsymbol{x}_{21}, ..., \boldsymbol{x}_{2n_2}]$ at random. We define a cross data matrix by $\boldsymbol{S}_{D(1)} = \{(n_1 - 1)(n_2 - 1)\}^{-1/2}(\boldsymbol{X}_1 - \overline{\boldsymbol{X}}_1)^T(\boldsymbol{X}_2 - \overline{\boldsymbol{X}}_2)$, where $\overline{\boldsymbol{X}}_i = [\bar{\boldsymbol{x}}_i, ..., \bar{\boldsymbol{x}}_i]^T$ is having p-vector $\bar{\boldsymbol{x}}_i = {n_i}^{-1} \sum_{j=1}^{n_i} \boldsymbol{x}_{ij}$ $(i = 1, 2)$. Let $r = n_2 - 1$. When we consider the singular value decomposition of $\boldsymbol{S}_{D(1)}$, it follows that $\boldsymbol{S}_{D(1)} = \sum_{j=1}^{r} \acute{\lambda}_j \acute{\boldsymbol{u}}_{j(1)} \acute{\boldsymbol{u}}_{j(2)}^T$, where $\acute{\lambda}_1 \geq \cdots \geq \acute{\lambda}_r (\geq 0)$ denote singular values of $\boldsymbol{S}_{D(1)}$, and $\acute{\boldsymbol{u}}_{j(1)}$ (or $\acute{\boldsymbol{u}}_{j(2)}$) denotes a unit left- (or right-) singular vector corresponding to $\acute{\lambda}_j$ $(j = 1, ..., r)$. Yata and Aoshima (2010, 2013) showed that $\acute{\lambda}_j$ has several consistency properties when $p \to \infty$ and $n \to \infty$. Again, we would like to emphasize that in this paper we focus on the largest eigenvalue and give its asymptotic properties when $p \to \infty$ while n is fixed.

Let us write $\boldsymbol{X}_i - [\boldsymbol{\mu}, ..., \boldsymbol{\mu}] = \boldsymbol{H}\boldsymbol{\Lambda}^{1/2}\boldsymbol{Z}_i$, where $\boldsymbol{Z}_i = [\boldsymbol{z}_{i1}, ..., \boldsymbol{z}_{ip}]^T$ and $\boldsymbol{z}_{ij} = (z_{ij1}, ..., z_{ijn_i})^T$, $i = 1, 2$; $j = 1, ..., p$. Let $\boldsymbol{z}_{oij} = \boldsymbol{z}_{ij} - (\bar{z}_{ij}, ..., \bar{z}_{ij})^T$, $j = 1, ..., p$, where $\bar{z}_{ij} = n_i^{-1} \sum_{k=1}^{n_i} z_{ijk}$ $(i = 1, 2; j = 1, ..., p)$. Then, we have that

$$\sqrt{(n_1 - 1)(n_2 - 1)}\boldsymbol{S}_{D(1)} = \lambda_1 \boldsymbol{z}_{o11}\boldsymbol{z}_{o21}^T + \sum_{j=2}^{p} \lambda_j \boldsymbol{z}_{o1j}\boldsymbol{z}_{o2j}^T.$$

Here, under (A-i), for any (i, j) element of $\sum_{j=2}^{p} \lambda_j \boldsymbol{z}_{o1j}\boldsymbol{z}_{o2j}^T$, it holds that as $p \to \infty$

$$\frac{\mathrm{Var}\{\sum_{s=1}^{p} \lambda_j (z_{1si} - \bar{z}_{1s})(z_{2sj} - \bar{z}_{2s})\}}{\lambda_1^2} = (1 - 1/n_1)(1 - 1/n_2)\frac{\mathrm{tr}(\boldsymbol{\Sigma}^2) - \lambda_1^2}{\lambda_1^2} \to 0.$$

Then, under (A-i) without (A-ii), we claim that as $p \to \infty$

$$\frac{\sum_{j=2}^{p} \lambda_j \boldsymbol{z}_{o1j}\boldsymbol{z}_{o2j}^T}{\lambda_1} \xrightarrow{P} \boldsymbol{O}.$$

Therefore, we have that

$$\frac{\acute{\lambda}_1}{\lambda_1} = \acute{\boldsymbol{u}}_{1(1)}^T \frac{\boldsymbol{S}_{D(1)}}{\lambda_1} \acute{\boldsymbol{u}}_{1(2)} = (\acute{\boldsymbol{u}}_{1(1)}^T \boldsymbol{z}_{o11}/\sqrt{n_1 - 1})(\boldsymbol{z}_{o21}^T \acute{\boldsymbol{u}}_{1(2)}/\sqrt{n_2 - 1}) + o_p(1) \tag{2.2}$$

under (A-i). Then, from (2.2), we have the following results.

Theorem 2.2. *Assume* $P(\lim_{p\to\infty} ||\boldsymbol{z}_{oi1}|| \neq 0) = 1$, $i = 1, 2$. *Under (A-i), it holds that as* $p \to \infty$

$$\frac{\acute{\lambda}_1}{\lambda_1} = ||\boldsymbol{z}_{o11}/\sqrt{n_1 - 1}||\,||\boldsymbol{z}_{o21}/\sqrt{n_2 - 1}|| + o_p(1).$$

Corollary 2.3. *If* z_{1j}, $j = 1, ..., n$, *are i.i.d. as* $N(0, 1)$, *it holds that as* $p \to \infty$

$$\frac{\acute{\lambda}_1}{\lambda_1} \Rightarrow \sqrt{\frac{\chi^2_{(1)n_1-1}}{n_1 - 1}}\sqrt{\frac{\chi^2_{(2)n_2-1}}{n_2 - 1}}$$

under (A-i), where $\chi^2_{(i)n_i-1}$ $(i = 1, 2)$ *denotes a random variable distributed as* χ^2 *distribution with* $n_i - 1$ *degrees of freedom, and* $\chi^2_{(1)n_1-1}$ *and* $\chi^2_{(2)n_2-1}$ *are independent.*

Now, we consider a bias correction of $\acute{\lambda}_1$. We have that

$$E\Big(\sqrt{\frac{\chi^2_{(1)n_1-1}}{n_1-1}}\sqrt{\frac{\chi^2_{(2)n_2-1}}{n_2-1}}\Big) = \frac{c}{\sqrt{(n_1-1)(n_2-1)}} \quad \text{and}$$

$$\mathrm{Var}\Big(\sqrt{\frac{\chi^2_{(1)n_1-1}}{n_1-1}}\sqrt{\frac{\chi^2_{(2)n_2-1}}{n_2-1}}\Big) = 1 - \frac{c^2}{(n_1-1)(n_2-1)}$$

$$\text{with} \quad c = 2\Gamma\Big(\frac{n_1}{2}\Big)\Gamma\Big(\frac{n_2}{2}\Big)\Gamma\Big(\frac{n_1-1}{2}\Big)^{-1}\Gamma\Big(\frac{n_2-1}{2}\Big)^{-1},$$

where $\Gamma(\cdot)$ is the gamma function. Therefore, we give a bias corrected CDM estimator by

$$\acute{\lambda}_{1*} = \frac{\sqrt{(n_1-1)(n_2-1)}}{c}\acute{\lambda}_1.$$

Then, we have the following result.

Corollary 2.4. *If $z_{1j}, j = 1, ..., n$, are i.i.d. as $N(0,1)$, it holds that as $p \to \infty$*

$$\frac{\acute{\lambda}_{1*}}{\lambda_1} \Rightarrow \frac{1}{c}\sqrt{\chi^2_{(1)n_1-1}}\sqrt{\chi^2_{(2)n_2-1}}$$

under (A-i).

Remark 2.3. We note that

$$E\Big(c^{-1}\sqrt{\chi^2_{(1)n_1-1}}\sqrt{\chi^2_{(2)n_2-1}}\Big) = 1 \quad \text{and} \quad \mathrm{Var}\Big(c^{-1}\sqrt{\chi^2_{(1)n_1-1}}\sqrt{\chi^2_{(2)n_2-1}}\Big) = \frac{(n_1-1)(n_2-1)}{c^2} - 1.$$

Also, note that

$$\mathrm{Var}\Big(c^{-1}\sqrt{\chi^2_{(1)n_1-1}}\sqrt{\chi^2_{(2)n_2-1}}\Big) > \mathrm{Var}\Big((n-1)^{-1}\chi^2_{n-1}\Big) = \frac{2}{n-1}.$$

Therefore, from Corollaries 2.1 and 2.4, we emphasize that $\acute{\lambda}_{1*}$ has the asymptotic distribution without (A-ii) when $p \to \infty$ while n is fixed, however, the asymptotic variance of $\acute{\lambda}_{1*}$ is larger than that of $\tilde{\lambda}_1$.

3 Application

In this section, we consider a confidence interval of the contribution ratio for the first principal component by using the NR estimator. Let a and b be constants satisfying $P\{a \le \chi^2_{n-1} \le b\} = 1 - \alpha$. Then, from Corollary 2.1, under (A-i) and (A-ii), if $z_{1j}, j = 1, ..., n$, are i.i.d. as $N(0,1)$, it holds that as $p \to \infty$

$$P\Big(\frac{\lambda_1}{\mathrm{tr}(\boldsymbol{\Sigma})} \in \Big[\frac{(n-1)\tilde{\lambda}_1}{b\kappa + (n-1)\tilde{\lambda}_1}, \frac{(n-1)\tilde{\lambda}_1}{a\kappa + (n-1)\tilde{\lambda}_1}\Big]\Big)$$

$$= P\Big(a \le (n-1)\frac{\tilde{\lambda}_1}{\lambda_1} \le b\Big) = 1 - \alpha + o(1),$$

where $\kappa = \mathrm{tr}(\boldsymbol{\Sigma}) - \lambda_1 = \sum_{s=2}^{p}\lambda_s$. From Lemma A.1 in Appendix, we give a consistent estimator of κ by $\hat{\kappa} = (n-1)(\mathrm{tr}(\boldsymbol{S}_D) - \tilde{\lambda}_1)/(n-2)$. Then, we have the following result.

Theorem 3.1. *Assume* $\liminf_{p\to\infty} \kappa/\lambda_1 > 0$. *Under (A-i) and (A-ii), if* $z_{1j}, j = 1, ..., n$, *are i.i.d. as* $N(0,1)$, *it holds that as* $p \to \infty$

$$P\Big(\frac{\lambda_1}{tr(\boldsymbol{\Sigma})} \in \Big[\frac{(n-1)\tilde{\lambda}_1}{b\hat{\kappa} + (n-1)\tilde{\lambda}_1}, \frac{(n-1)\tilde{\lambda}_1}{a\hat{\kappa} + (n-1)\tilde{\lambda}_1}\Big]\Big) = 1 - \alpha + o(1). \tag{3.1}$$

Remark 3.1. If $\kappa/\lambda_1 \to 0$ as $p \to \infty$, the contribution ratio of the first principal component is asymptotically 1 in the sense that $\lambda_1/\mathrm{tr}(\boldsymbol{\Sigma}) \to 1$ as $p \to \infty$. We emphasize that the conventional estimator, $\hat{\lambda}_1$, cannot yield a confidence interval of the contribution ratio when $p \to \infty$ while n is fixed because the contribution ratio of the first principal component is asymptotically 1 under (A-iii).

Let us construct a confidence interval of the contribution ratio for the first principal component. We used a microarray data by Alon et al. (1999). The data set was composed of 40 ($= n_1$) colon tumor samples and 22 ($= n_2$) normal colon tissue samples. The samples were analyzed by the Affymetrix oligonucleotide array. They chose 2000 ($= p$) genes with highest minimal intensity across the samples. We constructed a 95% confidence interval by using the NR estimator. Here, the constants (a, b) were chosen for the confidence interval to have a minimum length. The results were summarized in Table 1.

Table 1. The confidence interval (CI) of the first contribution ratio, $\tilde{\lambda}_1$ and $\hat{\kappa}$ for a microarray data set with $p = 2000$.

	CI	$\tilde{\lambda}_1$	$\hat{\kappa}$
Colon ($n_1 = 40$)	[0.3601, 0.5797]	921	1078
Normal ($n_2 = 22$)	[0.3038, 0.5995]	867	1132

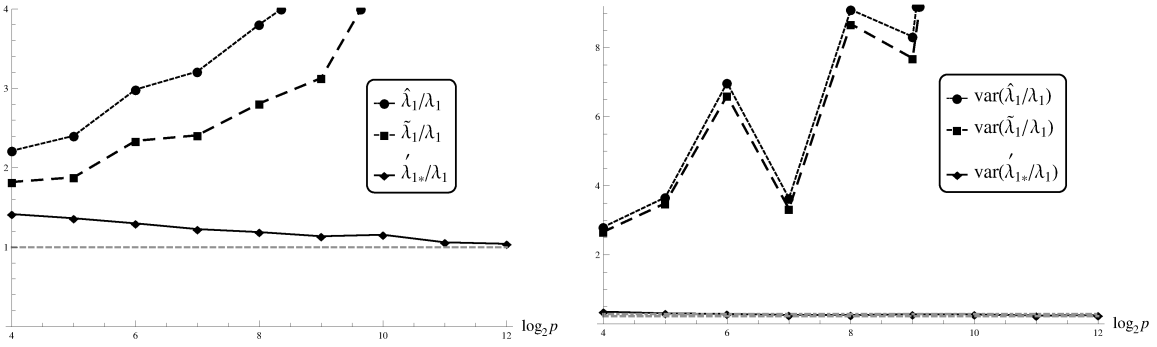
4 Simulation Studies

4.1 Comparisons of the Largest Eigenvalue Estimators

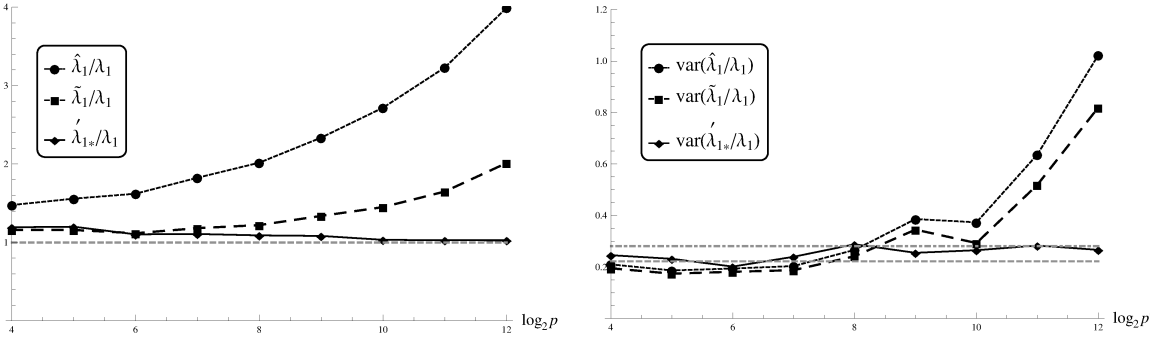
In order to compare the performances of the three largest eigenvalue estimators, $\hat{\lambda}_1$, $\tilde{\lambda}_1$ and $\acute{\lambda}_{1*}$, we used computer simulations. We set $\boldsymbol{\Sigma} = \mathrm{diag}(\lambda_1, ..., \lambda_p)$ with $\lambda_1 = p^{2/3}$, $\lambda_2 = p^{1/3}$ and $\lambda_3 = \cdots = \lambda_p = 1$. Note that (A-i) holds, however (A-iii) does not hold. We considered the cases of $p = 2^k (k = 4, ...12)$ and $n = 10$. We considered three distributions: (a) z_{1j}, $j = 1, ..., n$, are i.i.d. as $N(0,1)$ and $(z_{2j}, ..., z_{pj})^T$, $j = 1, ..., n$, are i.i.d. as a $(p-1)$-variate t-distribution, $t_{p-1}(\boldsymbol{0}, \boldsymbol{I}_{p-1}, 5)$, with mean zero vector, covariance matrix $\boldsymbol{I}_{p-1}$ and 5 degrees of freedom, where z_{1j} and $(z_{2j}, ..., z_{pj})^T$ are independent for each j; (b) z_{1j}, $j = 1, ..., n$, are i.i.d. as $N(0,1)$ and $(z_{2j}, ..., z_{pj})^T$, $j = 1, ..., n$, are i.i.d. as $t_{p-1}(\boldsymbol{0}, \boldsymbol{I}_{p-1}, 15)$, where z_{1j} and $(z_{2j}, ..., z_{pj})^T$ are independent for each j; and (c) $\boldsymbol{X}$ is Gaussian. Note that (A-ii) does not hold for (a) and (b). For (c), (A-ii) holds. Also, note that $t(\boldsymbol{0}, \boldsymbol{I}_p, \nu) \Rightarrow N_p(0, \boldsymbol{I}_p)$ as $\nu \to \infty$.

The findings were obtained by averaging the outcomes from 2000 ($= R$, say) replications. Under a fixed scenario, suppose that the r-th replication ends with estimates, $\hat{\lambda}_{1r}$, $\tilde{\lambda}_{1r}$ and $\acute{\lambda}_{1*r}$ for $r = 1, ..., R$. Let us simply write $\hat{\lambda}_1 = R^{-1}\sum_{r=1}^R \hat{\lambda}_{1r}$, $\tilde{\lambda}_1 = R^{-1}\sum_{r=1}^R \tilde{\lambda}_{1r}$ and $\acute{\lambda}_{1*} = R^{-1}\sum_{r=1}^R \acute{\lambda}_{1*r}$. We also considered the Monte Carlo variability. Let $\mathrm{var}(\hat{\lambda}_1/\lambda_1) = (R-1)^{-1}\sum_{r=1}^R (\hat{\lambda}_{1r} - \hat{\lambda}_1)^2/\lambda_1^2$, $\mathrm{var}(\tilde{\lambda}_1/\lambda_1) = (R-1)^{-1}\sum_{r=1}^R (\tilde{\lambda}_{1r} - \tilde{\lambda}_1)^2/\lambda_1^2$ and $\mathrm{var}(\acute{\lambda}_{1*}/\lambda_1) = (R-1)^{-1}\sum_{r=1}^R (\acute{\lambda}_{1*r} - \acute{\lambda}_{1*})^2/\lambda_1^2$. Then, we gave $\hat{\lambda}_1/\lambda_1$, $\tilde{\lambda}_1/\lambda_1$ and $\acute{\lambda}_{1*}/\lambda_1$ for $p = 2^k (k = 4, ...10)$ in the left panel of Fig. 1. In the right panel of Fig. 1, we gave $\mathrm{var}(\hat{\lambda}_1/\lambda_1)$, $\mathrm{var}(\tilde{\lambda}_1/\lambda_1)$, and $\mathrm{var}(\acute{\lambda}_{1*}/\lambda_1)$.

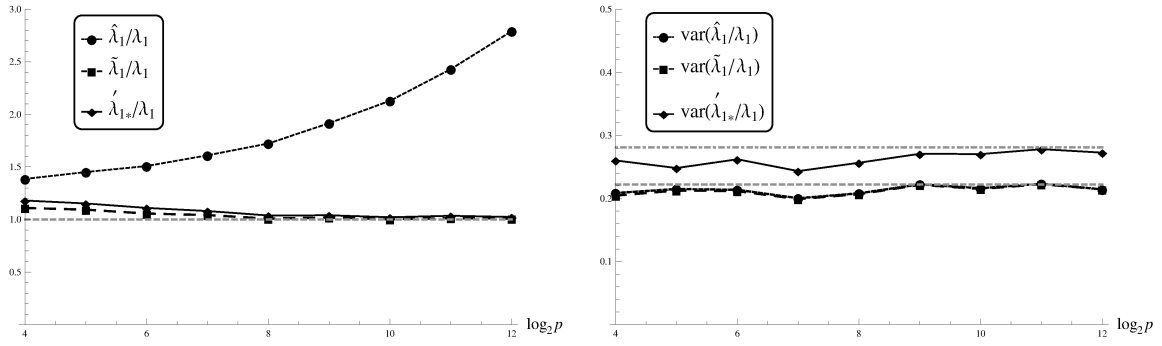
Throughout, the conventional estimator, $\hat{\lambda}_1$, gave bad performances. In case of (a), $\tilde{\lambda}_1$ by the NR method did not give always preferable performances especially when p is large. This is probably due to $\nu = 5$ that is not large enough for $\boldsymbol{X}$ to satisfy (A-ii). Contrary to that, $\acute{\lambda}_{1*}$ by the CDM method showed a quite good performance in (a). The NR method improved the performance in (b) and gave an excellent performance in (c) that satisfies (A-ii). We also observed that the NR method improves the Monte Carlo variability as well. Contrary to that, the CDM method does not improve the Monte Carlo variability. Thus, we conclude that if one can assume (A-ii), we recommend the NR method. Otherwise, one may use the CDM method freely from the assumption.



(a) $z_{1j},\ j = 1, ..., n$, are i.i.d. as $N(0, 1)$ and $(z_{2j}, ..., z_{pj})^T,\ j = 1, ..., n$, are i.i.d. as $t_{p-1}(\boldsymbol{0}, \boldsymbol{I}_{p-1}, 5)$.



(b) $z_{1j},\ j = 1, ..., n$, are i.i.d. as $N(0, 1)$ and $(z_{2j}, ..., z_{pj})^T,\ j = 1, ..., n$, are i.i.d. as $t_{p-1}(\boldsymbol{0}, \boldsymbol{I}_{p-1}, 15)$.



(c) When $\mathbf{X}$ is Gaussian.

Figure 1. The values of $\hat{\lambda}_1/\lambda_1$, $\tilde{\lambda}_1/\lambda_1$ and $\acute{\lambda}_{1*}/\lambda_1$ for $p = 2^k (k = 4, ...12)$ in the left panel. The values of $\mathrm{var}(\hat{\lambda}_1/\lambda_1)$, $\mathrm{var}(\tilde{\lambda}_1/\lambda_1)$, and $\mathrm{var}(\acute{\lambda}_{1*}/\lambda_1)$ for $p = 2^k (k = 4, ...12)$ in the right panel. In the right panel, the dashed lines denote $\mathrm{var}((n-1)^{-1}\chi^2_{n-1}) = 0.222$ and the chain lines denote $\mathrm{var}(c^{-1}\sqrt{\chi^2_{(1)n_1-1}}\sqrt{\chi^2_{(2)n_2-1}}) = 0.281$.

4.2 Confidence Interval of the First Contribution Ratio

In order to study the performance of the confidence interval of the contribution ratio for the first principal component by (3.1), we used computer simulations. Our goal was to construct a 95% confidence interval by (3.1), so we set $\alpha = 0.05, a = \chi^2_{n-1}(0.975)$ and $b = \chi^2_{n-1}(0.025)$, where $\chi^2_\nu(\beta)$ denotes the upper β point of χ^2_ν. We considered the cases of $p = 20, 100, 500$ and 2500 when $n = 10$. We set $\mathbf{\Sigma} = \mathrm{diag}(\lambda_1, ..., \lambda_p)$ with $\lambda_1 = p^{2/3}$ and $\lambda_2 = \cdots = \lambda_p = 1$. We considered $\boldsymbol{x}_j$, $j = 1, ..., n$, as z_{1j} being distributed as $N(0,1)$ and z_{ij}, $i = 2, ..., p$, being i.i.d. as $t_{p-1}(\mathbf{0}, \boldsymbol{I}_{p-1}, 5)$, where z_{1j} and $(z_{2j}, ..., z_{pj})$ are independent. Note that (A-i) and (A-ii) hold, however (A-iii) does not hold.

Independent pseudorandom 2000 (= R, say) observations of $\tilde{\lambda}_1$ and $\hat{\kappa}$ were generated from the distribution. Let $\tilde{\lambda}_{1r}$ and $\hat{\kappa}_r$ be the r-th observation of $\tilde{\lambda}_1$ and $\hat{\kappa}$ respectively, for $r = 1, ..., R$. Let us simply write $\tilde{\lambda}_1 = R^{-1}\sum_{r=1}^R \tilde{\lambda}_{1r}$ and $\hat{\kappa} = R^{-1}\sum_{r=1}^R \hat{\kappa}_r$. We also considered the Monte Carlo variability. Let $\mathrm{var}(\tilde{\lambda}_1/\lambda_1) = (R-1)^{-1}\sum_{r=1}^R (\tilde{\lambda}_{1r} - \tilde{\lambda}_1)^2/\lambda_1^2$ and $\mathrm{var}(\hat{\kappa}/\kappa) = (R-1)^{-1}\sum_{r=1}^R (\hat{\kappa}_r - \hat{\kappa})^2/\kappa^2$. In the end of the rth replication, we checked whether $\lambda_1/\mathrm{tr}(\mathbf{\Sigma})$ does (or does not) belong to the corresponding confidence interval and defined $P_r = 1$ (or 0) accordingly. Let $\overline{P}(0.95) = R^{-1}\sum_{r=1}^R P_r$, which estimates the target coverage probability, having its estimated standard error $s\{\overline{P}(0.95)\}$, where $s^2\{\overline{P}(0.95)\} = R^{-1}\overline{P}(0.95)(1 - \overline{P}(0.95))$. In Table 2, we gave $\overline{P}(0.95)$, $s\{\overline{P}(0.95)\}$, $\tilde{\lambda}_1/\lambda_1$, $\mathrm{var}(\tilde{\lambda}_1/\lambda_1)$, $\hat{\kappa}/\kappa$ and $\mathrm{var}(\hat{\kappa}/\kappa)$. We observed from Table 2 that $\overline{P}(0.95)$s become close to 0.95 as p increases. In addition, $\mathrm{var}(\tilde{\lambda}_1/\lambda_1)$s become close to $\mathrm{Var}(\chi^2_{n-1}/(n-1)) = 2/(n-1) \approx 0.222$ as p increases.

Table 2. The coverage probability of the first contribution ratio, $\bar{P}(0.95)$, together with $\tilde{\lambda}_1/\lambda_1$, $\hat{\kappa}/\kappa$ and their standard errors in parentheses.

p	$\bar{P}(0.95)$ $(s\{\overline{P}(0.95)\})$	$\tilde{\lambda}_1/\lambda_1$ $(\mathrm{var}(\tilde{\lambda}_1/\lambda_1))$	$\hat{\kappa}/\kappa$ $(\mathrm{var}(\hat{\kappa}/\kappa))$
20	0.961 (0.00430)	1.032 (0.192)	0.973 (0.00245)
100	0.963 (0.00419)	1.053 (0.218)	0.993 (0.00113)
500	0.963 (0.00422)	1.025 (0.214)	0.997 (0.00050)
2500	0.957 (0.00453)	1.018 (0.221)	0.999 (0.00022)

A Appendix

The following lemma was given by Ishii et al. (2014).

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

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

Proofs of Theorem 2.2, Corollaries 2.3 and 2.4. Let $\mathbf{1}_n = (1, ..., 1)^T \in \boldsymbol{R}^n$. We note that $\acute{\boldsymbol{u}}_{1(i)}^T \mathbf{1}_{n_i} = 0$, $i = 1, 2$, with probability tending to 1 under $P(\lim_{p\to\infty} ||\boldsymbol{z}_{oi1}|| \neq 0) = 1$, $i = 1, 2$. Also, note that $\boldsymbol{z}_{oi1}^T \mathbf{1}_{n_i} = 0$, $i = 1, 2$. Thus from (2.2), we have that $\acute{\boldsymbol{u}}_{1(i)} \xrightarrow{P} (\boldsymbol{z}_{oi1}/\sqrt{n_i - 1})/||\boldsymbol{z}_{oi1}/\sqrt{n_i - 1}||$, $i = 1, 2$, so that it concludes the result of Theorem 2.2. Note that $||\boldsymbol{z}_{oi1}||^2 = \sum_{k=1}^{n_i} z_{i1k}^2 - n_i \bar{z}_{i1}^2$ is distributed as $\chi_{n_1-1}^2$ for $i = 1, 2$, if z_{1j}, $j = 1, ..., k$, are i.i.d. as $N(0, 1)$. Thus we can conclude the results of Corollaries 2.3 and 2.4. □

Proof of Theorem 3.1. Under $\liminf_{p\to\infty} \kappa/\lambda_1 > 0$, from Lemma A.1, it holds that as $p \to \infty$ $\hat{\kappa}/\kappa = 1 + o_p(1)$. From Corollary 2.1, it holds under (A-i) and (A-ii) that

$$P\Big(\frac{\lambda_1}{\mathrm{tr}(\boldsymbol{\Sigma})} \in \Big[\frac{(n-1)\tilde{\lambda}_1}{b\hat{\kappa} + (n-1)\tilde{\lambda}_1}, \frac{(n-1)\tilde{\lambda}_1}{a\hat{\kappa} + (n-1)\tilde{\lambda}_1}\Big]\Big) = P\Big(\frac{(n-1)\tilde{\lambda}_1}{b\hat{\kappa} + (n-1)\tilde{\lambda}_1} \leq \frac{\lambda_1}{\mathrm{tr}(\boldsymbol{\Sigma})} \leq \frac{(n-1)\tilde{\lambda}_1}{a\hat{\kappa} + (n-1)\tilde{\lambda}_1}\Big)$$

$$= P\Big(\frac{a\hat{\kappa}}{(n-1)\tilde{\lambda}_1} \leq \frac{\kappa}{\lambda_1} \leq \frac{b\hat{\kappa}}{(n-1)\tilde{\lambda}_1}\Big) = P\Big(a \leq (n-1)\frac{\tilde{\lambda}_1}{\lambda_1} \leq b\Big) + o(1) = 1 - \alpha + o(1).$$

It concludes the result. □

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 26800078. 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 26540010.

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.

[2] Alon, U., Barkai, N., Notterman, D. A., Gish, K., Ybarra, S., Mack, D., and Levine, A. J. (1999). Broad Patterns of Gene Expression Revealed by Clustering Analysis of Tumor and Normal Colon Tissues Probed by Oligonucleotide Arrays, *Proceedings of the National Academy of Sciences of the United States of America* 96: 6745-6750.

[3] Aoshima, M. and Yata, K. (2011a). Two-Stage Procedures for High-Dimensional Data, *Sequantial Analysis (Editor's special invited paper)* 30: 356-399.

[4] Aoshima, M. and Yata, K. (2011b). Authors' Response, *Sequantial Analysis* 30: 432-440.

[5] 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.

[6] Aoshima, M. and Yata, K. (2013b). Invited Review Article: Statistical Inference for High-Dimension, Low-Sample-Size Data, *Sugaku* 65: 225-247.

[7] Aoshima, M. and Yata, K. (2013c). The JSS Research Prize Lecture: Effective Methodologies for High-Dimensional Data, *Journal of the Japan Statistical Society, Series J* 43: 123-150.

[8] Hall, P., Marron, J.S., and Neeman, A. (2005). Geometric Representation of High Dimension, Low Sample Size Data, *Journal of Royal Statistical Society, Series B* 67: 427-444.

[9] Ishii, A., Yata, K., and Aoshima, M. (2014). Asymptotic Distribution of the Largest Eigenvalue via Geometric Representations of High-Dimension, Low-Sample-Size Data, *Sri Lankan Journal of Applied Statistics*, Special Issue: Modern Statistical Methodologies in the Cutting Edge of Science (ed. Mukhopadhyay, N.), in press.

[10] Jung, S. and Marron, J.S. (2009). PCA Consistency in High Dimension, Low Sample Size Context, *Annals of Statistics* 37: 4104-4130.

[11] 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.

[12] 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.

[13] 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.

[14] Yata, K. and Aoshima, M. (2013). PCA Consistency for the Power Spiked Model in High-Dimensional Settings, *Journal of Multivariate Analysis* 122: 334-354.

Institute of Mathematics, University of Tsukuba
Ibaraki 305-8571, Japan
E-mail address: aoshima@math.tsukuba.ac.jp

筑波大学・数理物質科学研究科　石井　晶
筑波大学・数理物質系　矢田和善
筑波大学・数理物質系　青嶋　誠