Title	Asymptotic Properties of the First Principal Component (New Advances in Statistical Inference and Its Related Topics)
Author(s)	石井, 晶
Citation	数理解析研究所講究録 (2015), 1954: 44-50
Issue Date	2015-06
URL	http://hdl.handle.net/2433/224022
Right	
Type	Departmental Bulletin Paper
Textversion	publisher

数理解析研究所講究録
第 1954 巻 2015 年 44-50

Asymptotic Properties of the First Principal Component

筑波大学・数理物質科学研究科　石井　晶 (Aki Ishii)
Graduate School of Pure and Applied Sciences
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 the first principal component (PC) by HDLSS asymptotics in which the sample size is fixed when the data dimension grows. We use the noise-reduction (NR) methodology to estimate the first PC in HDLSS situations. We show that the eigenvalue estimator by the NR method holds preferable asymptotic properties under mild conditions when the data dimension is high. We provide an asymptotic distribution of the NR eigenvalue estimator in the HDLSS asymptotics. We also give asymptotic properties of the first PC direction and PC score in the HDLSS asymptotics. Finally, we summarize simulation results.

Keywords: 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). However, Yata and Aoshima (2009) developed HDLSS theory without assuming those assumptions and showed that the naive 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 and discussed the 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 a 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 $\boldsymbol{\Lambda}_p$ =diag$(\lambda_{1(p)}, ..., \lambda_{p(p)})$ having 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 $\text{Var}(\boldsymbol{z}_{j(p)}) = \boldsymbol{I}_n$, 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 λ_1 has multiplicity one in the sense that $\liminf_{p\to\infty} \lambda_1/\lambda_2 > 1$. Also, we assume that $\limsup_{p\to\infty} E(z_{ij}^4) < \infty$ for all i, j and $P(\lim_{p\to\infty} ||\boldsymbol{z}_1|| \neq 0) = 1$. As necessary, we consider the following assumption for z_{1j}, $j = 1, ..., n$:

(A-i) z_{1j}, $j = 1, ..., n$, are i.i.d. as $N(0, 1)$.

Note that $P(\lim_{p\to\infty} ||\boldsymbol{z}_1|| \neq 0) = 1$ under (A-i). 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 eigendecomposition of $\boldsymbol{S}_D$ as $\boldsymbol{S}_D = \sum_{j=1}^{n-1} \hat{\lambda}_j \hat{\boldsymbol{u}}_j \hat{\boldsymbol{u}}_j^T$, where $\hat{\boldsymbol{u}}_j = (\hat{u}_{j1}, ..., \hat{u}_{jn})^T$ denotes a unit eigenvector corresponding to $\hat{\lambda}_j$. Note that $\boldsymbol{S}$ and $\boldsymbol{S}_D$ share non-zero eigenvalues.

In this paper, we study the first PC by HDLSS asymptotics in which $p \to \infty$ while n is fixed. In Section 2, we show that the eigenvalue estimator by the NR method holds preferable asymptotic properties under mild conditions when the data dimension is high. We provide an asymptotic distribution of the NR eigenvalue estimator in the HDLSS asymptotics. In Section 3, we also give asymptotic properties of the first PC direction and PC score in the HDLSS asymptotics. 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 in the HDLSS asymptotics. Let $\delta_i = \text{tr}(\boldsymbol{\Sigma}^2) - \sum_{s=1}^i \lambda_s^2 = \sum_{s=i+1}^p \lambda_s^2$ for $i = 1, ..., p-1$. We consider the following assumptions for the largest eigenvalue:

(A-ii) $\dfrac{\delta_1}{\lambda_1^2} = o(1)$ as $p \to \infty$ when n is fixed; $\dfrac{\delta_{i_*}}{\lambda_1^2} = o(1)$ as $p \to \infty$ for some fixed i_* $(< p)$ when $n \to \infty$.

(A-iii) $\dfrac{\sum_{r,s\geq 2}^p \lambda_r \lambda_s E\{(z_{rk}^2 - 1)(z_{sk}^2 - 1)\}}{n\lambda_1^2} = o(1)$ as $p \to \infty$ either when n is fixed or $n \to \infty$.

Note that (A-iii) holds when $\boldsymbol{X}$ is Gaussian and (A-ii) is met. Let $\boldsymbol{z}_{oj} = \boldsymbol{z}_j - (\bar{z}_j, ..., \bar{z}_j)^T$, $j = 1, ..., p$, where $\bar{z}_j = n^{-1}\sum_{k=1}^n z_{jk}$. Let $\kappa = \text{tr}(\boldsymbol{\Sigma}) - \lambda_1 = \sum_{s=2}^p \lambda_s$. Then, we have the following result.

Proposition 2.1 (Ishii et al., 2015). *Under (A-ii) and (A-iii), it holds that*

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

as $p \to \infty$ either when n is fixed or $n \to \infty$.

Remark 2.1. Jung et al. (2012) gave a result similar to Proposition 2.1 when $\boldsymbol{X}$ is Gaussian, $\boldsymbol{\mu} = \boldsymbol{0}$ and n is fixed.

It holds that $E(||z_{o1}/\sqrt{n-1}||^2) = 1$ and $||z_{o1}/\sqrt{n-1}||^2 = 1 + o_p(1)$ as $n \to \infty$. If $\kappa/(n\lambda_1) = o(1)$ as $p \to \infty$ and $n \to \infty$, $\hat{\lambda}_1$ is a consistent estimator of λ_1. When n is fixed, the condition '$\kappa/\lambda_1 = o(1)$' is equivalent to '$\lambda_1/\text{tr}(\Sigma) = 1+o(1)$' in which the contribution ratio of the first principal component is asymptotically 1. In that sense, '$\kappa/\lambda_1 = o(1)$' is a quite strict condition for real high-dimensional data. Hereafter, we assume $\liminf_{p\to\infty} \kappa/\lambda_1 > 0$.

Yata and Aoshima (2012) proposed a method for eigenvalue estimation called the *noise-reduction (NR) methodology* that was brought by a geometric representation of S_D. If one applies the NR methodology to the present case, λ_is are estimated by

$$\tilde{\lambda}_i = \hat{\lambda}_i - \frac{\text{tr}(S_D) - \sum_{j=1}^{i} \hat{\lambda}_j}{n-1-i} \quad (i = 1, ..., n-2). \tag{2.1}$$

Note that $\tilde{\lambda}_i \geq 0$ w.p.1 for $i = 1, ..., n-2$. Also, note that the second term in (2.1) with $i = 1$ is an estimator of $\kappa/(n-1)$. Yata and Aoshima (2012, 2013) showed that $\tilde{\lambda}_i$ has several consistency properties when $p \to \infty$ and $n \to \infty$. On the other hand, Ishii et al. (2014) gave asymptotic properties of $\tilde{\lambda}_1$ when $p \to \infty$ while n is fixed. The following theorem summarizes their findings:

Theorem 2.1 (Ishii et al., 2015). *Under (A-ii) and (A-iii), it holds that as $p \to \infty$*

$$\frac{\tilde{\lambda}_1}{\lambda_1} = \begin{cases} ||z_{o1}/\sqrt{n-1}||^2 + o_p(1) & \text{when } n \text{ is fixed,} \\ 1 + o_p(1) & \text{when } n \to \infty. \end{cases}$$

Under (A-i) to (A-iii), it holds that as $p \to \infty$

$$(n-1)\frac{\tilde{\lambda}_1}{\lambda_1} \Rightarrow \chi^2_{n-1} \quad \text{when } n \text{ is fixed,}$$

$$\sqrt{\frac{n-1}{2}}\Big(\frac{\tilde{\lambda}_1}{\lambda_1} - 1\Big) \Rightarrow N(0,1) \quad \text{when } n \to \infty.$$

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.

3 Asymptotic Properties of the First PC Direction and PC Score

In this section, we consider asymptotic properties of the first PC direction and PC score in the HDLSS asymptotics.

3.1 First PC direction

Let $\hat{H} = [\hat{h}_1, ..., \hat{h}_p]$, where $\hat{H}$ is a $p \times p$ orthogonal matrix of the sample eigenvectors such that $\hat{H}^T S \hat{H} = \hat{\Lambda}$ having $\hat{\Lambda} = \text{diag}(\hat{\lambda}_1, ..., \hat{\lambda}_p)$. We assume $h_i^T \hat{h}_i \geq 0$ w.p.1 for all i without loss of generality. Note that $\hat{h}_i$ can be calculated by $\hat{h}_i = \{(n-1)\hat{\lambda}_i\}^{-1/2}(X - \overline{X})\hat{u}_i$. First, we have the following result.

Lemma 3.1. *Under (A-ii) and (A-iii), it holds that*

$$\hat{h}_1^T h_1 = \Big(1 + \frac{\kappa}{\lambda_1 ||z_{o1}||^2}\Big)^{-1/2} + o_p(1)$$

as $p \to \infty$ either when n is fixed or $n \to \infty$.

If $\kappa/(n\lambda_1) = o(1)$ as $p \to \infty$ and $n \to \infty$, $\hat{\boldsymbol{h}}_1$ is a consistent estimator of $\boldsymbol{h}_1$ in the sense that $\hat{\boldsymbol{h}}_1^T \boldsymbol{h}_1 = 1 + o_p(1)$. When n is fixed, $\hat{\boldsymbol{h}}_1$ is not a consistent estimator because $\lim_{p\to\infty} \kappa/\lambda_1 > 0$. In order to overcome this inconvenience, we consider applying the NR methodology to the PC direction vector. Let $\tilde{\boldsymbol{h}}_i = \{(n-1)\tilde{\lambda}_i\}^{-1/2}(\boldsymbol{X} - \overline{\boldsymbol{X}})\hat{\boldsymbol{u}}_i$. From Lemma 3.1, we have the following result.

Theorem 3.1. *Under (A-ii) and (A-iii), it holds that*

$$\tilde{\boldsymbol{h}}_1^T \boldsymbol{h}_1 = 1 + o_p(1)$$

as $p \to \infty$ either when n is fixed or $n \to \infty$.

Note that $||\tilde{\boldsymbol{h}}_1||^2 = \hat{\lambda}_1/\tilde{\lambda}_1 \geq 1$ w.p.1. One can claim that $\tilde{\boldsymbol{h}}_1$ is a consistent estimator of $\boldsymbol{h}_1$ in the sense of the inner product even when n is fixed though $\tilde{\boldsymbol{h}}_1$ is not a unit vector.

3.2 First PC score

We note that the first PC score is given by $s_{1j} = \sqrt{\lambda_1} z_{o1j}$ $(j = 1, ..., n)$, where $\boldsymbol{z}_{o1} = (z_{o11}, ..., z_{o1n})^T$. We apply the NR method to estimate the first PC score by $\tilde{s}_{1j} = \sqrt{(n-1)\tilde{\lambda}_1}\hat{u}_{1j}$ $(j = 1, ..., n)$, where $\hat{\boldsymbol{u}}_1 = (\hat{u}_{11}, ..., \hat{u}_{1n})^T$. Then, we have the following result.

Theorem 3.2. *Under (A-ii) and (A-iii), it holds that as $p \to \infty$*

$$\frac{1}{\sqrt{\lambda_1}}\tilde{s}_{1j} = \frac{1}{\sqrt{\lambda_1}}s_{1j} + o_p(1),\ j = 1, ..., n.$$

Remark 3.1. The naive estimator of the first PC score is given by $\hat{s}_{1j} = \sqrt{(n-1)\hat{\lambda}_1}\hat{u}_{1j}$ $(j = 1, ..., n)$. Under (A-ii) and (A-iii), it holds that as $p \to \infty$

$$\frac{1}{\sqrt{\lambda_1}}\hat{s}_{1j} = \frac{1}{\sqrt{\lambda_1}}\left(1 + \frac{\kappa}{\lambda_1||\boldsymbol{z}_{o1}||}\right)^{-1/2} s_{1j} + o_p(1),\ j = 1, ..., n$$

when n is fixed. If $\kappa/(n\lambda_1) = o(1)$ as $p \to \infty$ and $n \to \infty$, $\hat{s}_{1j}$ is a consistent estimator of s_{1j}.

4 Simulation Studies

In this section, we compare the performances of $\tilde{\lambda}_1$, $\tilde{\boldsymbol{h}}_1$ and $\tilde{s}_{1j}$ with naive estimators by Monte Carlo simulations. We set $p = 2^k$, $k = 3, ..., 11$ and $n = 10$. We considered two cases for λ_is: (I) $\lambda_i = p^{1/i}$, $i = 1, ..., p$ and (II) $\lambda_i = p^{4/(2+3i)}$, $i = 1, ..., p$. Note that $\lambda_1 = p$ for (I) and $\lambda_1 = p^{4/5}$ for (II). Also, note that (A-ii) holds both for (I) and (II). Let $p_* = \lceil p^{1/2} \rceil$, where $\lceil x \rceil$ denotes the smallest integer $\geq x$. We considered a non-Gaussian distribution as follows: $(z_{1j}, ..., z_{p-p_*j})^T$, $j = 1, ..., n$, are i.i.d. as $N_{p-p_*}(\boldsymbol{0}, \boldsymbol{I}_{p-p_*})$ and $(z_{p-p_*+1j}, ..., z_{pj})^T$, $j = 1, ..., n$, are i.i.d. as p_*-variate t-distribution, $t_{p_*}(\boldsymbol{0}, \boldsymbol{I}_{p_*}, 15)$, with mean zero, covariance matrix $\boldsymbol{I}_{p_*}$ and degrees of freedom 15, where $(z_{1j}, ..., z_{p-p_*j})^T$ and $(z_{p-p_*+1j}, ..., z_{pj})^T$ are independent for each j. Note that (A-i) and (A-iii) hold both for (I) and (II) from the fact that

$$\sum_{r,s\geq 2}^{p} \lambda_r\lambda_s E\{(z_{rk}^2 - 1)(z_{sk}^2 - 1)\} = 2\sum_{s=2}^{p-p_*} \lambda_s^2 + O(\sum_{r,s\geq p-p_*+1}^{p} \lambda_r\lambda_s) = o(\lambda_1^2).$$

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, $(\hat{\lambda}_{1r}, \hat{\boldsymbol{h}}_{1r}, \mathrm{MSE}(\hat{s}_1)_r)$ and $(\tilde{\lambda}_{1r}, \tilde{\boldsymbol{h}}_{1r}, \mathrm{MSE}(\tilde{s}_1)_r)$ $(r = 1, ..., R)$, where $\mathrm{MSE}(\hat{s}_1)_r = n^{-1}\sum_{j=1}^{n}(\hat{s}_{1j(r)} - s_{1j})^2$ and $\mathrm{MSE}(\tilde{s}_1)_r = n^{-1}\sum_{j=1}^{n}(\tilde{s}_{1j(r)} - s_{1j})^2$. Let us simply write $\hat{\lambda}_1 = R^{-1}\sum_{r=1}^{R}\hat{\lambda}_{1r}$ and $\tilde{\lambda}_1 = R^{-1}\sum_{r=1}^{R}\tilde{\lambda}_{1r}$. We also considered the Monte Carlo variability by $\mathrm{var}(\hat{\lambda}_1/\lambda_1) = (R-1)^{-1}\sum_{r=1}^{R}(\hat{\lambda}_{1r} - \hat{\lambda}_1)^2/\lambda_1^2$ and $\mathrm{var}(\tilde{\lambda}_1/\lambda_1) = (R-1)^{-1}\sum_{r=1}^{R}(\tilde{\lambda}_{1r} - \tilde{\lambda}_1)^2/\lambda_1^2$. Figure 1 shows the behaviors of $(\hat{\lambda}_1/\lambda_1, \tilde{\lambda}_1/\lambda_1)$ for (I) and (II). Figure 2 shows the behaviors of $(\mathrm{var}(\hat{\lambda}_1/\lambda_1), \mathrm{var}(\tilde{\lambda}_1/\lambda_1))$ for (I) and (II). We gave the asymptotic variance of $\tilde{\lambda}_1/\lambda_1$ by $\mathrm{Var}\{\chi_{n-1}^2/(n-1)\} = 0.222$ from Theorem 2.1 and showed it by the solid line in Figure 2. We observed that the sample mean and variance of $\tilde{\lambda}_1/\lambda_1$ become close to those asymptotic values as p increases.

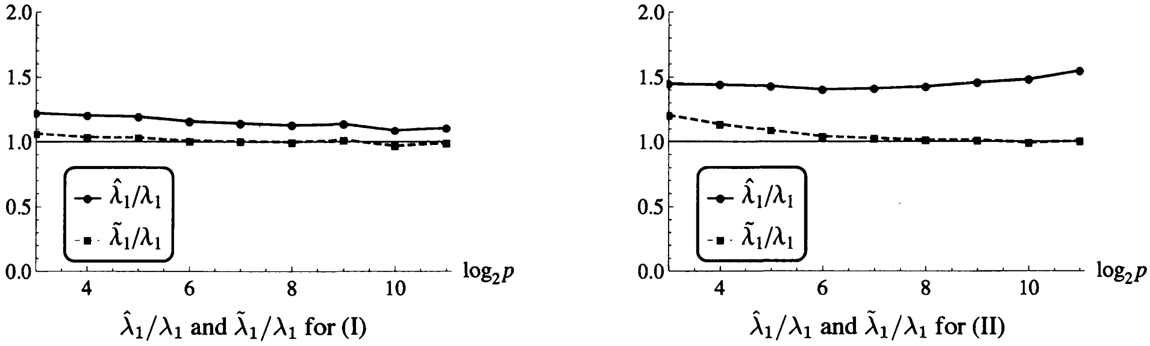


Figure 1. The values of $\hat{\lambda}_1/\lambda_1$ is denoted by the solid line and $\tilde{\lambda}_1/\lambda_1$ is denoted by the dashed line for (I) in the left panel and for (II) in the right panel.

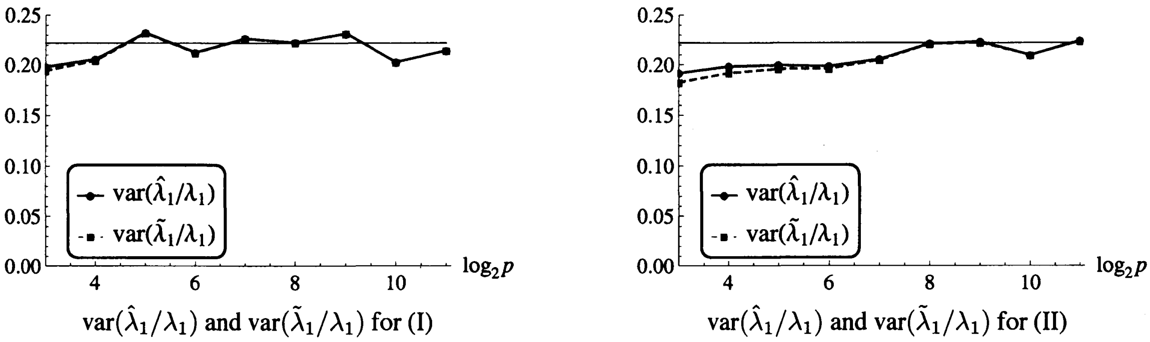
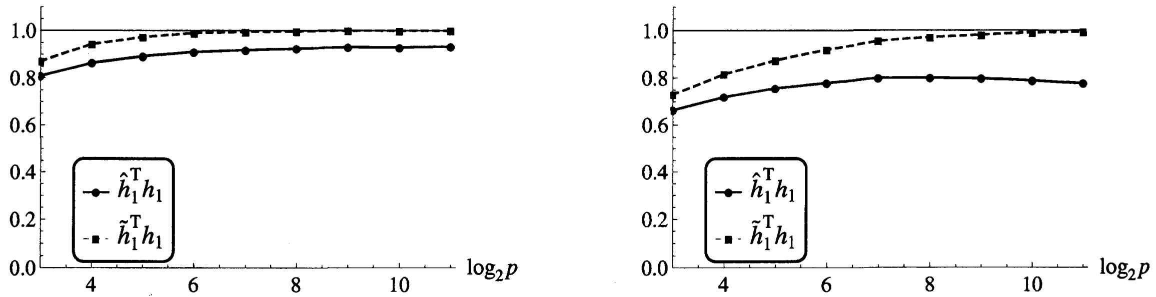


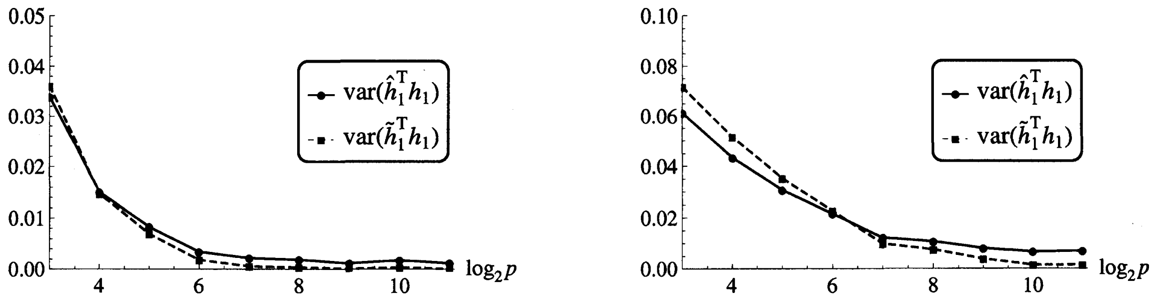
Figure 2. The values of $\mathrm{var}(\hat{\lambda}_1/\lambda_1)$ is denoted by solid line and $\mathrm{var}(\tilde{\lambda}_1/\lambda_1)$ is denoted by dashed line for (I) in the left panel and for (II) in the right panel. The asymptotic variance was given by $\mathrm{Var}\{\chi_{n-1}^2/(n-1)\} = 0.222$ and denoted by the solid line.

Similarly, we plotted $(\hat{\boldsymbol{h}}_1^T\boldsymbol{h}_1, \tilde{\boldsymbol{h}}_1^T\boldsymbol{h}_1)$, $(\mathrm{var}(\hat{\boldsymbol{h}}_1^T\boldsymbol{h}_1), \mathrm{var}(\tilde{\boldsymbol{h}}_1^T\boldsymbol{h}_1))$ and $(\mathrm{MSE}(\hat{s}_1)/\lambda_1, \mathrm{MSE}(\tilde{s}_1)/\lambda_1)$ in Figure 3, Figure 4 and Figure 5. Throughout, the estimators by the NR method gave good performances both for (I) and (II) when p is large. However, the naive estimators gave poor performances especially for (II). This is probably because the bias of the naive estimators, $\kappa/(n\lambda_1)$, is large for (II) compared to (I). See Proposition 2.1 for the details.



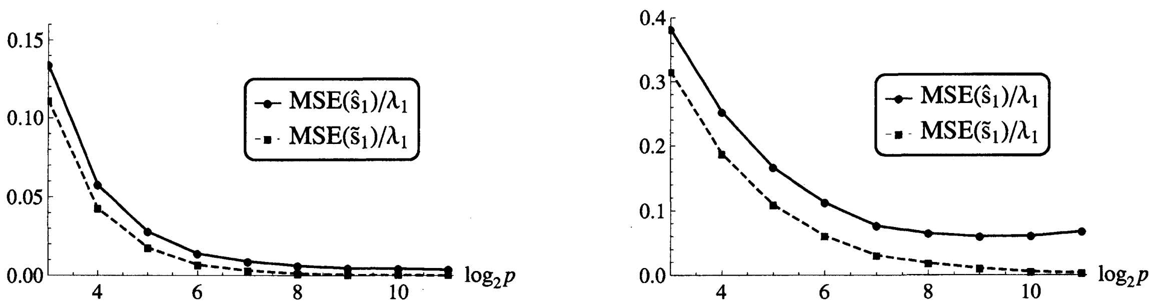
$\hat{\boldsymbol{h}}_1^T \boldsymbol{h}_1$ and $\tilde{\boldsymbol{h}}_1^T \boldsymbol{h}_1$ for (I) $\hat{\boldsymbol{h}}_1^T \boldsymbol{h}_1$ and $\tilde{\boldsymbol{h}}_1^T \boldsymbol{h}_1$ for (II)

Figure 3. The values of $\hat{\boldsymbol{h}}_1^T \boldsymbol{h}_1$ is denoted by the solid line and $\tilde{\boldsymbol{h}}_1^T \boldsymbol{h}_1$ is denoted by the dashed line for (I) in the left panel and for (II) in the right panel.



$\mathrm{var}(\hat{\boldsymbol{h}}_1^T \boldsymbol{h}_1)$ and $\mathrm{var}(\tilde{\boldsymbol{h}}_1^T \boldsymbol{h}_1)$ for (I) $\mathrm{var}(\hat{\boldsymbol{h}}_1^T \boldsymbol{h}_1)$ and $\mathrm{var}(\tilde{\boldsymbol{h}}_1^T \boldsymbol{h}_1)$ for (II)

Figure 4. The values of $\mathrm{var}(\hat{\boldsymbol{h}}_1^T \boldsymbol{h}_1)$ is denoted by the solid line and $\mathrm{var}(\tilde{\boldsymbol{h}}_1^T \boldsymbol{h}_1)$ is denoted by the dashed line for (I) in the left panel and for (II) in the right panel.



$\mathrm{MSE}(\hat{s}_1)/\lambda_1$ and $\mathrm{MSE}(\tilde{s}_1)/\lambda_1$ for (I) $\mathrm{MSE}(\hat{s}_1)/\lambda_1$ and $\mathrm{MSE}(\tilde{s}_1)/\lambda_1$ for (II)

Figure 5. The values of $\mathrm{MSE}(\hat{s}_1)/\lambda_1$ is denoted by the solid line and $\mathrm{MSE}(\tilde{s}_1)/\lambda_1$ is denoted by the dashed line for (I) in the left panel and for (II) in the right panel.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Professor Makoto Aoshima, University of Tsukuba, for his enthusiastic guidance and helpful support to my research project.

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

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

[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 for High-Dimension, Low-Sample-Size Data, *Sugaku* 65: 225-247.

[6] 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.

[7] 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.

[8] 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.), 81-94.

[9] Ishii, A., Yata, K., and Aoshima, M. (2015). Asymptotic Properties of the First Principal Component and Equality Tests of Covariance Matrices in High-Dimension, Low-Sample-Size Context, submitted. arXiv:1503.07302.

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

[11] Jung, S., Sen A., and Marron, J.S. (2012). Boundary Behavior in High Dimension, Low Sample Size Asymptotics of PCA, *Journal of Multivariate Analysis* 109: 190-203.

[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.

[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.

[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.

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