Tests of high-dimensional mean vectors under the SSE model

Aki Ishii\textsuperscript{a}, Kazuyoshi Yata\textsuperscript{b} and Makoto Aoshima\textsuperscript{b}

\textsuperscript{a} Department of Information Sciences, Tokyo University of Science
\textsuperscript{b} Institute of Mathematics, University of Tsukuba

Abstract

In this paper, we discuss inference problems on high-dimensional mean vectors under the strongly spiked eigenvalue (SSE) model. First, we consider one-sample test. In order to avoid huge noise, we derive a new test statistic by using a data transformation technique. We show that the asymptotic normality can be established for the new test statistic. We give an asymptotic size and power of a new test procedure.

\textit{Key words and phrases:} Asymptotic normality; Data transformation; Eigenstructure estimation; Large \( p \) small \( n \); Noise reduction methodology; Spiked model

1 Introduction

In this paper, we consider statistical inference on mean vectors in the high-dimension, low-sample-size (HDLSS) context. Let \( x_1, \ldots, x_n \) be a random sample of size \( n \geq 4 \) from a \( p \)-variate distribution with an unknown mean vector \( \mu \) and unknown covariance matrix \( \Sigma \). In the HDLSS context, the data dimension \( p \) is very high and \( n \) is much smaller than \( p \). We define the eigen-decomposition of \( \Sigma \) by \( \Sigma = HH^T \), where \( \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p) \) is a diagonal matrix of eigenvalues, \( \lambda_1 \geq \cdots \geq \lambda_p \geq 0 \), and \( H = [h_1, \ldots, h_p] \) is an orthogonal matrix of the corresponding eigenvectors. We write the sample mean vector and the sample covariance matrix as \( \bar{x} = \sum_{j=1}^{n} x_j / n \) and \( S = \sum_{j=1}^{n} (x_j - \bar{x})(x_j - \bar{x})^T / (n - 1) \).

In this paper, we discuss the one-sample test:

\[ H_0 : \ \mu = \mu_0 \quad \text{vs.} \quad H_1 : \ \mu \neq \mu_0, \] (1)

where \( \mu_0 \) is a candidate mean vector. We assume \( \mu_0 = 0 \) without loss of generality. One should note that Hotelling’s \( T^2 \)-statistic defined by

\[ T^2 = n^{-1} \bar{x}^T S^{-1} \bar{x} \]

is not available because \( S^{-1} \) does not exist in the HDLSS context. Dempster (1958, 1960) considered the test when \( X \) is Gaussian. When \( X \) is non-Gaussian, Bai and Saranadasa (1996) considered the test. They considered a test statistic defined by

\[ T_{\text{BS}} = \| \bar{x} \|^2 - \text{tr}(S) / n, \] (2)
where \( \| \cdot \| \) denotes the Euclidean norm. Let \( \Delta = \| \mu \|^2 \). Note that \( E(T_{\text{ns}}) = \Delta \) and \( \text{Var}(T_{\text{ns}}) = K \), where \( K = K_1 + K_2 \) having
\[
K_1 = 2 \frac{\text{tr}(\Sigma^2)}{n(n-1)} \quad \text{and} \quad K_2 = 4 \frac{\mu^T \Sigma \mu}{n}.
\]

Let us consider the following eigenvalue condition:
\[
\frac{\lambda_i^2}{\text{tr}(\Sigma^2)} \to 0 \quad \text{as} \quad p \to \infty. \tag{3}
\]

Let
\[
m = \min\{p, n\}.
\]

Under (3), \( H_0 \) and some regularity conditions, Bai and Saranadasa (1996) and Aoshima and Yata (2011, 2015) showed the asymptotic normality as follows:
\[
T_{\text{ns}} / K_1^{1/2} \Rightarrow N(0, 1) \quad \text{as} \quad m \to \infty. \tag{4}
\]

Here, “\( \Rightarrow \)” denotes the convergence in distribution and \( N(0, 1) \) denotes a random variable distributed as the standard normal distribution.

Aoshima and Yata (2018a) called the eigenvalue condition (3) the “non-strongly spiked eigenvalue (NSSE) model” and drew attention that high-dimensional data do not fit the NSSE model on several occasions. In order to overcome this inconvenience, Aoshima and Yata (2018a) proposed the “strongly spiked eigenvalue (SSE) model” defined by
\[
\lim \inf_{p \to \infty} \left\{ \frac{\lambda_i^2}{\text{tr}(\Sigma^2)} \right\} > 0 \tag{5}
\]
and gave a data transformation technique from the SSE model to the NSSE model.

**Remark 1.1.** When we consider a spiked model such as
\[
\lambda_r = a_r  \rho^{\alpha_r} \quad (r = 1, \ldots, t) \quad \text{and} \quad \lambda_r = c_r \quad (r = t + 1, \ldots, p) \tag{6}
\]
with positive and fixed constants, \( a_r, s, c_r, s \) and \( \alpha_r, s \), and a positive and fixed integer \( t \). Note that the NSSE condition (3) is met when \( \alpha_1 < 1/2 \). On the other hand, the SSE condition (5) is met when \( \alpha_1 \geq 1/2 \). We emphasize that high-dimensional data often have the SSE model. For instance, when we analyze a microarray data, we find several gene networks in which genes in the same network are highly correlated each other. The high correlation is one of the reasons why strongly spiked eigenvalues appear in high-dimensional data analysis.

Let us show a toy example about the asymptotic null distribution of \( T_{\text{ns}} \) in (4). We set \( p = 2000 \) and \( n = 90 \). We assumed that \( X \) is Gaussian as \( N_p(0, \Sigma) \) under \( H_0 : \mu = 0 \). We set \( \Sigma = \text{diag}(\rho^{\alpha_1}, \rho^{\alpha_2}, 1, \ldots, 1) \) and considered two cases:
(i) \((\alpha_1, \alpha_2) = (1/3, 1/4)\) and (ii) \((\alpha_1, \alpha_2) = (2/3, 1/2)\).
When \((1; 2) = (3; 1)\) \((i)\) and \((2; 3) = (1; 2)\) \((ii)\).

Figure 1: The histograms of \(T_{bs}/K_{1}^{1/2}\) for a NSSE model \((i)\) and for a SSE model \((ii)\) when \((p, n) = (2000, 90)\). The solid line denotes the p.d.f. of \(N(0, 1)\).

Note that \((i)\) is a NSSE model and \((ii)\) is a SSE model. We generated independent pseudo-random observations for each case and calculated \(T_{bs}/K_{1}^{1/2}\) 1000 times. In Fig.1, we gave histograms for \((i)\) and \((ii)\). One can observe that \(T_{bs}/K_{1}^{1/2}\) does not converge to \(N(0, 1)\) in case \((ii)\).

It is necessary to develop a new test statistic instead of \(T_{bs}\) for the SSE model. Katayama et al. (2013) gave a test statistic and showed that it has an asymptotic null distribution as a \(\chi^{2}\)-distribution when \(X\) is Gaussian. When \(X\) is non-Gaussian, Ishii et al. (2016) gave a different test statistic by using the noise-reduction (NR) method due to Yata and Aoshima (2012) and provided an asymptotic non-null distribution to discuss the size and power when \(p \to \infty\) while \(n\) is fixed. However, the performance of those test statistics is not always preferable because they are heavily influenced by strongly spiked eigenvalues and the variance of their statistics becomes very large because of the huge noise.

In this paper, we avoid the huge noise by using the data transformation technique and construct a new test procedure for the SSE model.

2 High-dimensional one-sample test by data transformation

2.1 Assumptions

We write the square root of \(M\) as \(M^{1/2}\) for any positive-semidefinite matrix \(M\). Let

\[
x_{l} = H \Lambda^{1/2} z_{l} + \mu,
\]

where \(z_{l} = (z_{1l}, ..., z_{pl})^{T}\) is considered as a sphered data vector having the zero mean vector and identity covariance matrix. Similar to Bai and Saranadasa (1996) and Chen and Qin (2010), we assume the following assumption as necessary:

\[
(A-i) \quad \limsup_{p \to \infty} E(z_{rl}^{4}) < \infty \text{ for all } r, \quad E(z_{rl}^{2}z_{sl}^{2}) = E(z_{rl}^{2})E(z_{sl}^{2}) = 1, \quad E(z_{rl}z_{sl}z_{tl}) = 0 \quad \text{and} \quad E(z_{rl}z_{sl}z_{tl}z_{ul}) = 0 \text{ for all } r \neq s, t, u.
\]
When $x_l$s are Gaussian, (A-i) naturally holds. Let
\[
\Psi_r = \text{tr}(\Sigma^2) - \sum_{s=1}^{r-1} \lambda_s^2 = \sum_{s=r}^{p} \lambda_s^2 \quad \text{for } r = 1, \ldots, p.
\]
Similar to Aoshima and Yata (2018a), we assume the following model:

(A-ii) There exists a fixed integer $k (\geq 1)$ such that

(i) When $k \geq 2$, $\lambda_1, \ldots, \lambda_k$ are distinct in the sense that
\[
\liminf_{p \to \infty} (\lambda_r/\lambda_s - 1) > 0 \quad \text{for } 1 \leq r < s \leq k;
\]

(ii) $\lambda_k$ and $\lambda_{k+1}$ satisfy
\[
\liminf_{p \to \infty} \frac{\lambda_k^2}{\Psi_k} > 0 \quad \text{and} \quad \frac{\lambda_{k+1}^2}{\Psi_{k+1}} \to 0 \quad \text{as } p \to \infty.
\]

Note that (A-ii) is one of the SSE models. For example, (A-ii) holds when the spiked model in (6) has the constants such that
\[
\alpha_1 \geq \cdots \geq \alpha_k \geq 1/2 > \alpha_{k+1} \geq \cdots \geq \alpha_t \quad \text{and}
\]
\[
a_r \neq a_s \quad \text{for } 1 \leq r < s \leq k.
\]

### 2.2 Data transformation

According to Aoshima and Yata (2018a,b), we consider transforming the data from the SSE model to the NSSE model by using the projection matrix
\[
A = I_p - \sum_{j=1}^{k} h_j h_j^T = \sum_{j=k+1}^{p} h_j h_j^T.
\]
We have that $E(Ax_j) = A\mu$ (=$\mu_*$, say) and
\[
\text{Var}(Ax_j) = A\Sigma A = \sum_{j=k+1}^{p} \lambda_j h_j h_j^T (= \Sigma_*, \text{say}).
\]
Note that $\text{tr}(\Sigma_*^2) = \Psi_{k+1}$ and $\lambda_{\text{max}}(\Sigma_*) = \lambda_{k+1}$, where $\lambda_{\text{max}}(\Sigma_*)$ denotes the largest eigenvalue of $\Sigma_*$. Then, it holds that
\[
\lambda_{\text{max}}^2(\Sigma_*)/\text{tr}(\Sigma_*^2) \to 0 \quad \text{as } p \to \infty \text{ under (A-ii)}.
\]
Thus, the transformed data has the NSSE model.
By using the transformed data, we consider the following quantity:

\[
T_{dr} = \|A\bar{x}\|^2 - \frac{\text{tr}(AS)}{n} = 2\frac{\sum_{l<l'} x_l^T A x_{l'}}{n(n-1)} = 2\frac{\sum_{l<l'} \left(x_l^T x_{l'} - \sum_{j=1}^{k} x_{jl} x_{j'l'}\right)}{n(n-1)},
\]

where

\[
x_{jl} = h_j^T x_l \quad \text{for all } j, l.
\]

Let \(\Delta_* = \|\mu_*\|^2\). Note that \(E(T_{dr}) = \Delta_*\) and \(\text{Var}(T_{dr}) = K_*\), where \(K_* = K_{1*} + K_{2*}\) having

\[
K_{1*} = 2\frac{\text{tr}(\Sigma_*)}{n(n-1)} \quad \text{and} \quad K_{2*} = 4\frac{\mu_*^T \Sigma_* \mu_*}{n}.
\]

Under (A-ii), we consider the following conditions as necessary:

(A-iii) \(\lim \sup_{m \to \infty} \frac{\Delta_*^2}{K_{1*}} < \infty\); \quad (A-iv) \(\frac{K_{1*}}{\Delta_*^2} \to 0\) as \(m \to \infty\).

We note that (A-iii) is met under \(H_0\) in (1). Then, we have the following results.

**Proposition 2.1** (Ishii, Yata and Aoshima, 2018). Assume (A-i) to (A-iii). Then, it holds that as \(m \to \infty\)

\[
\frac{T_{dr} - \Delta_*}{K_{1*}^{1/2}} = \frac{T_{dr} - \Delta_*}{K_{1*}^{1/2}} + o_P(1) \Rightarrow N(0, 1).
\]

**Proposition 2.2** (Ishii, Yata and Aoshima, 2018). Assume (A-ii) and (A-iv). Then, it holds that as \(m \to \infty\)

\[
\frac{T_{dr}}{\Delta_*} = 1 + o_P(1).
\]

### 3 New test statistic and its asymptotic properties

Based on \(T_{dr}\), we define the test statistic as follows:

\[
\hat{T}_{dr} = 2\frac{\sum_{l<l'} \left(x_l^T x_{l'} - \sum_{j=1}^{k} \tilde{x}_{jl} \tilde{x}_{j'l'}\right)}{n(n-1)} = \|\bar{x}\|^2 - \text{tr}(S)/n - 2\sum_{l<l'} \sum_{j=1}^{k} \tilde{x}_{jl} \tilde{x}_{j'l'}/n(n-1),
\]

where \(\tilde{x}_{jl}\) is a certain estimator of \(x_{jl}\). We give the definition of \(\tilde{x}_{jl}\) in Appendix A since its derivation is somewhat technical.

We assume the following conditions when (A-ii) is met.

(A-v) \(\frac{\lambda_*^2}{n\text{tr}(\Sigma_*^2)} \to 0\) as \(m \to \infty\); \quad (A-vi) \(\lim \inf_{p \to \infty} \frac{\Delta_*}{\Delta} > 0\) when \(\Delta \neq 0\).
Note that (A-vi) is a mild condition when \( p \) is much larger than \( k \) because \( \Delta_\ast = \Delta - \sum_{j=1}^{k} (h_j^T \mu)^2 = \sum_{j=k+1}^{\nu} (h_j^T \mu)^2 \). Also, note that \( \Delta_\ast \neq 0 \) when \( \Delta \neq 0 \) under (A-vi). Then, we have the following results.

**Theorem 3.1** (Ishii, Yata and Aoshima, 2018). Assume (A-i), (A-ii), (A-v) and (A-vi). It holds that
\[
\sqrt{\frac{K_{1, s}^{1/2}}{K_{s}^{1/2}}} \left( \frac{\Delta_\ast}{\Delta} \right) \rightarrow \mathcal{N}(0, 1) \quad \text{as} \quad m \rightarrow \infty.
\]
Furthermore, we also assume (A-iii). Then, it holds that
\[
\sqrt{\frac{K_{1, s}^{1/2}}{K_{s}^{1/2}}} \left( \frac{\Delta_\ast}{\Delta} \right) \rightarrow \mathcal{N}(0, 1) \quad \text{as} \quad m \rightarrow \infty.
\]

**Corollary 3.1** (Ishii, Yata and Aoshima, 2018). Assume (A-i), (A-ii) and (A-iv) to (A-vi). Then, it holds that
\[
\sqrt{\frac{K_{1, s}^{1/2}}{K_{s}^{1/2}}} \left( \frac{\Delta_\ast}{\Delta} \right) \rightarrow \mathcal{N}(0, 1) \quad \text{as} \quad m \rightarrow \infty.
\]

From Theorem 3.1 and Lemma A.1, under (A-i) to (A-iii), (A-v) and (A-vi), it holds that
\[
\sqrt{\frac{K_{1, s}^{1/2}}{K_{s}^{1/2}}} \left( \frac{\Delta_\ast}{\Delta} \right) \rightarrow \mathcal{N}(0, 1) \quad \text{as} \quad m \rightarrow \infty,
\]
where \( \widehat{K}_{1, s} \) is given in Appendix A. Thus, we can construct a test procedure of (1) by (8).

We give a new test procedure for (1) by using \( \widehat{T}_{dr} \). For a given \( \alpha \in (0, 1/2) \), we test the hypothesis (1) by
\[
\text{rejecting } H_0 \iff \frac{\widehat{T}_{dr}}{\sqrt{K_{s}^{1/2}}} > z_{\alpha}, \tag{9}
\]
where \( z_{\alpha} \) is a constant such that \( P\{ \mathcal{N}(0, 1) > z_{\alpha} \} = \alpha \). From Theorem 3.1, Corollary 3.1 and Lemma A.1, we have the following results.

**Theorem 3.2** (Ishii, Yata and Aoshima, 2018). Assume (A-i), (A-ii), (A-v) and (A-vi). The test procedure (9) holds that as \( m \rightarrow \infty \)
\[
\text{Size} = \alpha + o(1) \quad \text{and} \quad \text{Power} = \Phi \left( \frac{\Delta_\ast}{\sqrt{K_{s}^{1/2}}} - z_{\alpha} \left( \frac{K_{1, s}^{1/2}}{K_{s}^{1/2}} \right) \right) + o(1),
\]
where \( \Phi(\cdot) \) denotes the c.d.f. of \( \mathcal{N}(0, 1) \).

**Corollary 3.2** (Ishii, Yata and Aoshima, 2018). Assume (A-i), (A-ii) and (A-iv) to (A-vi). Under \( H_1 \), the test procedure (9) holds that as \( m \rightarrow \infty \)
\[
\text{Power} = 1 + o(1).
\]

**Remark 3.1.** The number \( k \) in \( \frac{\widehat{T}_{dr}}{\sqrt{K_{s}^{1/2}}} \) should be determined before the data transformation. See Aoshima and Yata (2018a) about a choice of \( k \).

In this talk, we apply the findings to multi-sample problems under the SSE models.
Appendix A

In this section, we give estimators for the parameters in our new test statistic, \( \hat{T}_{\text{nr}} \), and discuss their asymptotic properties.

### A.1 Estimation of \( x_{jl} \)

Let \( \mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_n], \overline{\mathbf{X}} = [\overline{\mathbf{x}}, ..., \overline{\mathbf{x}}] \) and \( \mathbf{P}_n = I_n - 1_n 1_n^T/n \), where \( 1_n = (1, ..., 1)^T \). Recall \( \mathbf{S} \) is the sample covariance matrix. One can write that \( \mathbf{S} = (\mathbf{X} - \overline{\mathbf{X}})(\mathbf{X} - \overline{\mathbf{X}})^T/(n-1) = \mathbf{X} \mathbf{P}_n \mathbf{X}^T/(n-1) \). Let us write the eigen-decomposition of \( \mathbf{S} \) as \( \mathbf{S} = \sum_{j=1}^p \lambda_j \mathbf{h}_j \mathbf{h}_j^T \) having eigenvalues \( \lambda_1 \geq \cdots \geq \lambda_p \geq 0 \) and the corresponding \( p \)-dimensional unit eigenvectors \( \mathbf{h}_1, ..., \mathbf{h}_p \). We assume \( P(h_j^T \mathbf{h}_j \geq 0) = 1 \) for all \( j \) without loss of generality. We also define the following \( n \times n \) dual sample covariance matrix:

\[
\mathbf{S}_D = (n-1)^{-1} \mathbf{P}_n \mathbf{X}^T \mathbf{P}_n = (n-1)^{-1} (\mathbf{X} - \overline{\mathbf{X}})^T (\mathbf{X} - \overline{\mathbf{X}}).
\]

Note that \( \mathbf{S} \) and \( \mathbf{S}_D \) share non-zero eigenvalues. Let us write the eigen-decomposition of \( \mathbf{S}_D \) as \( \mathbf{S}_D = \sum_{j=1}^p \lambda_j \mathbf{u}_j \mathbf{u}_j^T \), where \( \mathbf{u}_j = (\hat{u}_j, ..., \hat{u}_j)^T \) denotes a \( n \)-dimensional unit eigenvector corresponding to \( \lambda_j \). In high-dimensional settings, we calculate \( \mathbf{h}_j \) by using \( \mathbf{u}_j \) as follows:

\[
\hat{\mathbf{h}}_j = \{(n-1)\hat{\lambda}_j\}^{-1/2} (\mathbf{X} - \overline{\mathbf{X}}) \mathbf{u}_j.
\]

Note that \( 1_n^T \mathbf{S}_D 1_n = 0 \), so that \( 1_n^T \mathbf{u}_j = \sum_{n=1}^n \hat{u}_{jl} = 0 \) when \( \hat{\lambda}_j > 0 \).

For high-dimensional data, the sample eigenvalues and eigenvectors get huge noise. See Jung and Marron (2009) for the details. In order to remove the huge noise, Yata and Aoshima (2012) focused on a geometric representation of \( \mathbf{S}_D \) and proposed the NR method. If one applies the NR method, the \( \lambda_j \)'s and \( \mathbf{h}_j \)’s are estimated by

\[
\hat{\lambda}_j = \hat{\lambda}_j - \frac{\text{tr}(\mathbf{S}_D) - \sum_{l=1}^j \hat{\lambda}_l}{n-1-j} \quad (j = 1, ..., n-2) \quad \text{and} \quad \hat{\mathbf{h}}_j = \{(n-1)\hat{\lambda}_j\}^{-1/2} (\mathbf{X} - \overline{\mathbf{X}}) \mathbf{u}_j \quad (j = 1, ..., n-2).
\]

Note that \( P(\hat{\lambda}_j \geq 0) = 1 \) for \( j = 1, ..., n-2 \). We emphasize that \( \hat{\lambda}_j \)'s and \( \hat{\mathbf{h}}_j \)'s have consistency properties under much milder conditions than \( \lambda_j \)'s and \( \mathbf{h}_j \)'s. However, for the estimation of \( x_{jl} = x_{jl}^T \mathbf{h}_j \), Aoshima and Yata (2018a) showed that \( x_{jl}^T \hat{\mathbf{h}}_j \) involves a huge bias and gave a modification for all \( j, l \) by

\[
\tilde{x}_{jl} = \left( \frac{n-1}{n-2} \right) \frac{(\mathbf{X} - \overline{\mathbf{X}}) \hat{\mathbf{u}}_{jl}}{(n-1) \hat{\lambda}_j} = \left( \frac{n-1}{n-2} \right) \frac{(\mathbf{X} - \overline{\mathbf{X}}) \mathbf{u}_{jl}}{\hat{\lambda}_j^{1/2}},
\]

where

\[
\hat{\mathbf{u}}_{jl} = (\hat{u}_{jl}, ..., \hat{u}_{jl-1}, -\hat{u}_{jl}/(n-1), \hat{u}_{jl+1}, ..., \hat{u}_n)^T.
\]

Note that \( \sum_{l=1}^n \hat{u}_{jl}/n_l = \{(n-2)/(n-1)\} \hat{\mathbf{u}}_j \) and \( \sum_{l=1}^n \hat{x}_{jl}/n_l = \hat{\mathbf{x}}_j \). Then, we estimate \( x_{jl} \) by

\[
\tilde{x}_{jl} = \left( \frac{n-1}{n-2} \right) \frac{(\mathbf{X} - \overline{\mathbf{X}}) \hat{\mathbf{u}}_{jl}}{(n-1) \hat{\lambda}_j} \quad \text{for all } j, l.
\]

7
A.2 Estimation of $K_{1*}$

We use the CDM method given by Yata and Aoshima (2010) to estimate $K_{1*}$. Let $n(1) = \lceil n/2 \rceil$ and $n(2) = n - n(1)$, where $\lceil x \rceil$ denotes the smallest integer $\geq x$. Let $X_1 = [x_1, \ldots, x_{n(1)}]$ and $X_2 = [x_{n(1)+1}, \ldots, x_n]$. We define

$$S_{D(1)} = \{(n(1) - 1)(n(2) - 1)\}^{-1/2}(X_1 - \overline{X_1})^T(X_2 - \overline{X_2}),$$

where $X_i = [\overline{x}_{n(i),1}, \ldots, \overline{x}_{n(i),n(i)}]$ with $\overline{x}_{n(i),1} = \sum_{l=1}^{n(1)} x_l/n(1)$ and $\overline{x}_{n(i),n(i)} = \sum_{l=n(1)+1}^{n} x_l/n(2)$. We estimate $\lambda_j$ by the $j$-th singular value, $\hat{\lambda}_j$, of $S_{D(1)}$, where

$$\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_{n(2)-1} \geq 0.$$

Yata and Aoshima (2010) showed that $\hat{\lambda}_j$ has several consistency properties for high-dimensional non-Gaussian data. We note that $E\{\text{tr}(S_{D(1)}S_{D(1)}^T)\} = \text{tr}(\Sigma^2)$. We estimate $\Psi_r$ by $\hat{\Psi}_1 = \text{tr}(S_{D(1)}S_{D(1)}^T)$ and

$$\hat{\Psi}_r = \text{tr}(S_{D(1)}S_{D(1)}^T) - \sum_{s=1}^{r-1} \hat{\lambda}_s^2 \quad \text{for} \ r = 2, \ldots, n(2) - 1. \quad (13)$$

Note that $P(\hat{\Psi}_r \geq 0) = 1$ for $r = 1, \ldots, n(2) - 1$. Then, Aoshima and Yata (2018a) gave the following result.

**Lemma A.1** (Aoshima and Yata, 2018a). Assume (A-i) and (A-ii). Then, it holds that $\hat{\Psi}_r/\Psi_r = 1 + o_P(1)$ as $m \to \infty$ for $r = 1, \ldots, k + 1$.

Thus we estimate $\text{tr}(\Sigma_r^2)$ by $\hat{\Psi}_{k+1}$. Let

$$\hat{K}_{1*} = 2\hat{\Psi}_{k+1}/\{n(n-1)\}.$$

Then, from Lemma A.1, under (A-i) and (A-ii), it holds that

$$\hat{K}_{1*}/K_{1*} = 1 + o_P(1) \quad \text{as} \ m \to \infty.$$

**Acknowledgements**

Research of the first author was partially supported by Grant-in-Aid for Young Scientists, Japan Society for the Promotion of Science (JSPS), under Contract Number 18K18015. Research of the second author was partially supported by Grant-in-Aid for Scientific Research (C), JSPS, under Contract Number 18K03409. Research of the third author was partially supported by Grants-in-Aid for Scientific Research (A) and Challenging Research (Exploratory), JSPS, under Contract Numbers 15H01678 and 17K19956.
References


