Asymptotic Normality for Inference on Multisample, High-Dimensional Mean Vectors Under Mild Conditions

Makoto Aoshima · Kazuyoshi Yata

Abstract In this paper, we consider the asymptotic normality for various inference problems on multisample and high-dimensional mean vectors. We verify that the asymptotic normality of concerned statistics is proved under mild conditions for high-dimensional data. We show that the asymptotic normality can be justified theoretically and numerically even for non-Gaussian data. We introduce the extended cross-data-matrix (ECDM) methodology to construct an unbiased estimator at a reasonable computational cost. With the help of the asymptotic normality, we show that the concerned statistics given by ECDM can ensure consistency properties for inference on multisample and high-dimensional mean vectors. We give several applications such as confidence regions for high-dimensional mean vectors, confidence intervals for the squared norm and the test of multisample mean vectors. We also provide sample size determination so as to satisfy prespecified accuracy on inference. Finally, we give several examples by using a microarray data set.

1 Introduction

A common feature of high-dimensional data is that the data dimension is high, however, the sample size is relatively small. This is the so-called "HDLSS" or "large p, small n" situation where $p/n \to \infty$; here p is the data dimension and

---

M. Aoshima (✉) · K. Yata
Institute of Mathematics, University of Tsukuba, Ibaraki 305-8571, Japan
e-mail: aoshima@math.tsukuba.ac.jp

K. Yata
e-mail: yata@math.tsukuba.ac.jp

n is the sample size. The statistical inference on this type of data is becoming increasingly relevant. Suppose we have independent and p-variate populations, π_i, $i = 1, ..., k$, having unknown mean vector $\boldsymbol{\mu}_i$ and unknown covariance matrix $\boldsymbol{\Sigma}_i(\geq \boldsymbol{O})$ for each π_i. Let $\boldsymbol{\theta} = (\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_1, ..., \boldsymbol{\Sigma}_k)$. We do not assume either the normality of the population distributions or the equality of Σs as $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_k$. The eigen-decomposition of $\boldsymbol{\Sigma}_i$ $(i = 1, ..., k)$ is given by $\boldsymbol{\Sigma}_i = \boldsymbol{H}_i \boldsymbol{\Lambda}_i \boldsymbol{H}_i^T$, where $\boldsymbol{\Lambda}_i$ is a diagonal matrix of eigenvalues, $\lambda_{i1} \geq \cdots \geq \lambda_{ip} \geq 0$, and $\boldsymbol{H}_i$ is an orthogonal matrix of the corresponding eigenvectors. In this paper, we focus on inference on multisample and high-dimensional mean vectors. Let $\boldsymbol{\mu} = \sum_{i=1}^{k} b_i \boldsymbol{\mu}_i$, where b_is are known and nonzero scalars. Having recorded i.i.d. samples, $\boldsymbol{x}_{ij}$, $j = 1, ..., n_i$, of size n_i (≥ 4) from each π_i, we define $\boldsymbol{T}_{\mathbf{n}} = \sum_{i=1}^{k} b_i \bar{\boldsymbol{x}}_{in_i}$, where $\boldsymbol{n} = (n_1, ..., n_k)$ and $\bar{\boldsymbol{x}}_{in_i} = \sum_{j=1}^{n_i} \boldsymbol{x}_{ij}/n_i$.

Chen and Qin (2010) gave a two-sample test in inference on HDLSS data. Aoshima and Yata (2011a, b) developed a variety of inference on HDLSS data such as a given-bandwidth confidence region, a two-sample test, a test of equality of two covariance matrices, classification, variable selection, regression, pathway analysis and so on along with sample size determination for each inference. Yata and Aoshima (2012) provided given-width confidence intervals for the norm of mean vectors. Aoshima and Yata (2011a, b) and Yata and Aoshima (2012) assumed that $\boldsymbol{x}_{ij} = \boldsymbol{H}_i \boldsymbol{\Lambda}_i^{1/2} \boldsymbol{z}_{ij} + \boldsymbol{\mu}_i$ for $i = 1, ..., k$; $j = 1, ..., n_i$, where $\boldsymbol{z}_{ij} = (z_{i1j}, ..., z_{ipj})^T$, $E(\boldsymbol{z}_{ij}) = \boldsymbol{0}$, $\text{Var}(\boldsymbol{z}_{ij}) = \boldsymbol{I}_p$ and the fourth moments of each variable in $\boldsymbol{z}_{ij}$ are uniformly bounded. Here, $\boldsymbol{I}_p$ denotes the identity matrix of dimension p. They considered one of the following three assumptions for π_is:

(A-i) $\pi_i : N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ for $i = 1, ..., k$;

(A-ii) $z_{isj}, s = 1, ..., p$, are independent for $i = 1, ..., k$;

(A-iii) $E(z_{isj}^2 z_{itj}^2) = E(z_{isj}^2)E(z_{itj}^2)$ and $E(z_{isj} z_{itj} z_{iuj} z_{ivj}) = 0$ for $s \neq t, u, v$, and some regularity conditions given in Aoshima and Yata (2011a).

They assumed the following conditions for $\boldsymbol{\Sigma}_i$s:

(A-iv) $\liminf_{p\to\infty} \lambda_{ip} > 0$, $\limsup_{p\to\infty} \dfrac{\text{tr}(\boldsymbol{\Sigma}_i^t)}{p} < \infty$ $(t = 1, 2)$ and $\dfrac{\text{tr}(\boldsymbol{\Sigma}_i^4)}{p^2} \to 0$ as $p \to \infty$ for $i = 1, ..., k$.

Note that (A-i) implies (A-ii). It holds that $E_{\boldsymbol{\theta}}(\|\boldsymbol{T}_{\mathbf{n}} - \boldsymbol{\mu}\|^2) = \sum_{i=1}^{k} b_i^2 \text{tr}(\boldsymbol{\Sigma}_i)/n_i$ $(= \Sigma_{\mathbf{n}}$, say). Also, it holds that $\text{Var}_{\boldsymbol{\theta}}(\|\boldsymbol{T}_{\mathbf{n}} - \boldsymbol{\mu}\|^2) = 2\sum_{i,j} b_i^2 b_j^2 \text{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j)/(n_i n_j)$ under (A-i). Let $\boldsymbol{S}_{in_i} = \sum_{j=1}^{n_i}(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{in_i})(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{in_i})^T/(n_i - 1)$ and $\widehat{\Sigma}_{\mathbf{n}} = \sum_{i=1}^{k} b_i^2 \text{tr}(\boldsymbol{S}_{in_i})/n_i$. It holds that $E_{\boldsymbol{\theta}}(\|\boldsymbol{T}_{\mathbf{n}} - \boldsymbol{\mu}\|^2 - \widehat{\Sigma}_{\mathbf{n}}) = 0$ and

$$\text{Var}_{\boldsymbol{\theta}}(\|\boldsymbol{T}_{\mathbf{n}} - \boldsymbol{\mu}\|^2 - \widehat{\Sigma}_{\mathbf{n}}) = 2\sum_{i=1}^{k} \frac{b_i^4 \text{tr}(\boldsymbol{\Sigma}_i^2)}{n_i(n_i - 1)} + 4\sum_{i<j} \frac{b_i^2 b_j^2 \text{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j)}{n_i n_j} \quad (= K, \text{ say}).$$

Let $\widehat{T}_{\mathbf{n}} = \|\boldsymbol{T}_{\mathbf{n}}\|^2 - \widehat{\Sigma}_{\mathbf{n}}$. It holds that $E_{\boldsymbol{\theta}}(\widehat{T}_{\mathbf{n}}) = \|\boldsymbol{\mu}\|^2$ and $\text{Var}_{\boldsymbol{\theta}}(\widehat{T}_{\mathbf{n}}) = K + 4\sum_{i=1}^{k} b_i^2 \boldsymbol{\mu}^T \boldsymbol{\Sigma}_i \boldsymbol{\mu}/n_i$ $(= K_*, \text{ say})$. Then, the following results were obtained by Aoshima and Yata (2011a) and Yata and Aoshima (2012).

Theorem 1 (Aoshima and Yata 2011a) *Assume* (A-i) *and* (A-iv). *Then, it holds that*

$$\frac{\|\boldsymbol{T}_{\mathbf{n}} - \boldsymbol{\mu}\|^2 - \Sigma_{\mathbf{n}}}{\sqrt{2\sum_{i,j} b_i^2 b_j^2 \text{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j)/(n_i n_j)}} \Rightarrow N(0, 1) \qquad (1)$$

when $p \to \infty$ and either $n_i \to \infty$ or n_i is fixed for $i = 1, ..., k$, where "$\Rightarrow$" denotes the convergence in distribution and $N(0, 1)$ denotes a random variable distributed as the standard normal distribution.

Theorem 2 (Aoshima and Yata 2011a) *Assume* (A-iv) *and either* (A-ii) *or* (A-iii). *Then, Eq.* 1 *holds as $p \to \infty$ and $n_i \to \infty$, $i = 1, ..., k$.*

Corollary 1 (Aoshima and Yata 2011a) *Assume* (A-iv) *and either* (A-ii) *or* (A-iii). *Then, it holds as $p \to \infty$ and $n_i \to \infty$, $i = 1, ..., k$, that*

$$\frac{||\boldsymbol{T_n} - \boldsymbol{\mu}||^2 - \widehat{\Sigma}_\mathbf{n}}{K^{1/2}} \Rightarrow N(0, 1).$$

Theorem 3 (Yata and Aoshima 2012) *Assume that $n_i \boldsymbol{\mu}^T \boldsymbol{\Sigma}_i \boldsymbol{\mu} / tr(\boldsymbol{\Sigma}_i^2) = o(1)$, $i = 1, ..., k$. Assume* (A-iv) *and either* (A-ii) *or* (A-iii). *Then, it holds as $p \to \infty$ and $n_i \to \infty$, $i = 1, ..., k$, that*

$$\frac{\widehat{T}_\mathbf{n} - ||\boldsymbol{\mu}||^2}{K_*^{1/2}} \Rightarrow N(0, 1).$$

In this paper, we relax the conditions to verify the asymptotic normality of concerned statistics for high-dimensional data. Then, we apply the asymptotic normality to various inference problems on multisample and high-dimensional mean vectors. In Section 2, we verify that the asymptotic normality of concerned statistics is proved under mild conditions. We show that the asymptotic normality can be justified theoretically and numerically even for non-Gaussian data. In Section 3, we introduce the extended cross-data-matrix (ECDM) methodology, developed by Yata and Aoshima (2013), to construct an unbiased estimator at a reasonable computational cost. In Section 4, with the help of the asymptotic normality, we show that the statistics given by ECDM can ensure consistency properties for inference on multisample and high-dimensional mean vectors. We give several applications such as confidence regions for high-dimensional mean vectors, confidence intervals for the squared norm and the test of multisample mean vectors. In Section 5, we provide sample size determination so as to satisfy prespecified accuracy on inference. Finally, in Section 6, we give several examples by using a microarray data set.

2 Asymptotic normality under mild conditions

We assume that

$$\boldsymbol{x}_{ij} = \boldsymbol{\Gamma}_i \boldsymbol{w}_{ij} + \boldsymbol{\mu}_i \text{ for } i = 1, ..., k; \ j = 1, ..., n_i, \tag{2}$$

where $\boldsymbol{\Gamma}_i$ is a $p \times r_i$ matrix for some $r_i > 0$ such that $\boldsymbol{\Gamma}_i \boldsymbol{\Gamma}_i^T = \boldsymbol{\Sigma}_i$, and $\boldsymbol{w}_{ij}$, $j = 1, ..., n_i$, are i.i.d. random vectors having $E(\boldsymbol{w}_{ij}) = \mathbf{0}$ and $Var(\boldsymbol{w}_{ij}) = \boldsymbol{I}_{r_i}$. Note that Eq. 2 includes the case that $\boldsymbol{\Gamma}_i = \boldsymbol{H}_i \boldsymbol{\Lambda}_i^{1/2}$ and $\boldsymbol{w}_{ij} = \boldsymbol{z}_{ij}$. Refer to Bai and Sarandasa (1996), Chen and Qin (2010) and Aoshima and Yata (A distance-based, misclassification rate adjusted classifier for multiclass, high-dimensional data. To appear in Ann Inst

Stat Math) for the details of the model. Let $\boldsymbol{w}_{ij} = (w_{i1j}, ..., w_{ir_ij})^T$ for any i, j. As for $\boldsymbol{w}_{ij}$, we assume that

(A-v) The fourth moments of each variable in $\boldsymbol{w}_{ij}$ are uniformly bounded, $E(w_{isj}^2 w_{itj}^2) = E(w_{isj}^2) E(w_{itj}^2)$ and $E(w_{isj} w_{itj} w_{iuj} w_{ivj}) = 0$ for $s \neq t, u, v$.

Note that any of the assumptions, (A-i) to (A-iii), implies (A-v). That is (A-v) is milder than any of (A-i) to (A-iii). We assume the following conditions for $\boldsymbol{\Sigma}_i$s:

(A-vi) $\dfrac{\lambda_{i1}}{\text{tr}(\boldsymbol{\Sigma}_i^2)^{1/2}} \to 0$ as $p \to \infty$ for $i = 1, ..., k$.

We recall that λ_{i1} $(i = 1, ..., k)$ is the largest eigenvalue of $\boldsymbol{\Sigma}_i$. Note that $\{\lambda_{i1}/\text{tr}(\boldsymbol{\Sigma}_i^2)^{1/2}\}^4 \leq \text{tr}(\boldsymbol{\Sigma}_i^4)/\text{tr}(\boldsymbol{\Sigma}_i^2)^2$. We also note that $\liminf_{p \to \infty} \text{tr}(\boldsymbol{\Sigma}_i^2)/p > 0$ under the assumption that $\liminf_{p \to \infty} \lambda_{ip} > 0$. Thus (A-iv) implies (A-vi). That is (A-vi) is milder than (A-iv). Under (A-v) and (A-vi), we have the following result.

Theorem 4 *Assume* (A-v) *and* (A-vi). *Then, it holds as* $p \to \infty$ *and* $n_i \to \infty$, $i = 1, ..., k$, *that*

$$\frac{||\boldsymbol{T}_\mathbf{n} - \boldsymbol{\mu}||^2 - \widehat{\Sigma}_\mathbf{n}}{K^{1/2}} \Rightarrow N(0, 1).$$

We assume the following extra condition:

(A-vii) $\dfrac{\sum_{i=1}^k \boldsymbol{\mu}^T \boldsymbol{\Sigma}_i \boldsymbol{\mu}/n_i}{K} = o(1)$.

Note that the condition that $n_i \boldsymbol{\mu}^T \boldsymbol{\Sigma}_i \boldsymbol{\mu}/\text{tr}(\boldsymbol{\Sigma}_i^2) = o(1)$, $i = 1, ..., k$, in Theorem 3 implies (A-vii). Then, we have the following result.

Theorem 5 *Assume* (A-v) *to* (A-vii). *Then, it holds as* $p \to \infty$ *and* $n_i \to \infty$, $i = 1, ..., k$, *that*

$$\frac{\widehat{T}_\mathbf{n} - ||\boldsymbol{\mu}||^2}{K_*^{1/2}} \Rightarrow N(0, 1).$$

When (A-vii) is not met, we have the following result.

Corollary 2 *Assume* (A-vi) *and*

(A-viii) $\liminf \dfrac{\sum_{i=1}^k \boldsymbol{\mu}^T \boldsymbol{\Sigma}_i \boldsymbol{\mu}/n_i}{K} > 0$ *as* $p \to \infty$ *and* $n_i \to \infty$, $i = 1, ..., k$.

Then, it holds as $p \to \infty$ *and* $n_i \to \infty$, $i = 1, ..., k$, *that*

$$\frac{\widehat{T}_\mathbf{n}}{||\boldsymbol{\mu}||^2} = 1 + o_p(1).$$

We consider the following conditions instead of (A-v):

(A-v') $\text{Var}_\theta[\{(\boldsymbol{x}_{il} - \boldsymbol{\mu}_i)^T (\boldsymbol{x}_{jl'} - \boldsymbol{\mu}_j)\}^2] = O\{\text{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j)^2\}$ and $\text{Var}_\theta\{(\boldsymbol{x}_{il} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_j (\boldsymbol{x}_{il} - \boldsymbol{\mu}_i)\} = O\{\text{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j \boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j)\}$ as $p \to \infty$ for $i, j = 1, ..., k$.

Then, we have the following result.

Corollary 3 *After replacing* (A-v) *with* (A-v')*, the results in Theorems* 4 *and* 5 *are still justified.*

Remark 1 From Eqs. 14 and 16 in Appendix, we note that (A-v) implies (A-v').

Since $\boldsymbol{\Sigma}_i$s are unknown, it is necessary to estimate K (or K_*). Let us consider an estimator of K by

$$\widehat{K} = 2\sum_{i=1}^{k} \frac{b_i^4 W_{in_i}}{n_i(n_i-1)} + 4\sum_{i<j} \frac{b_i^2 b_j^2 \text{tr}(\boldsymbol{S}_{in_i}\boldsymbol{S}_{jn_j})}{n_i n_j},$$

where W_{in_i} is defined by Eq. 4. From Lemma 3 in Appendix, we note that $\widehat{K}/K = 1 + o_p(1)$ as $p \to \infty$ and $n_i \to \infty$, $i = 1, ..., k$, under (A-v). Then, we have the following result.

Corollary 4 *Assume* (A-v) *and* (A-vi)*. Then, it holds as* $p \to \infty$ *and* $n_i \to \infty$*,* $i = 1, ..., k$*, that*

$$\frac{||\boldsymbol{T_n} - \boldsymbol{\mu}||^2 - \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{n}}}{\widehat{K}^{1/2}} \Rightarrow N(0, 1).$$

Note that $K_*/K \to 1$ as $p \to \infty$ and $n_i \to \infty$, $i = 1, ..., k$, under (A-vii). Then, we have the following result.

Corollary 5 *Assume* (A-v) *to* (A-vii)*. Then, it holds as* $p \to \infty$ *and* $n_i \to \infty$*,* $i = 1, ..., k$*, that*

$$\frac{\widehat{T}_{\boldsymbol{n}} - ||\boldsymbol{\mu}||^2}{\widehat{K}^{1/2}} \Rightarrow N(0, 1).$$

Remark 2 After replacing (A-v) with (A-v'), the results in Corollaries 4 and 5 are still justified.

Let us observe Corollaries 4 and 5. We set $k = 2$, $b_1 = 1$, $b_2 = -1$, $n_1 = n_2 = 20$, $\boldsymbol{\Sigma}_1 = \boldsymbol{B}(0.3^{|i-j|^{1/3}})\boldsymbol{B}$ and $\boldsymbol{\Sigma}_2 = \boldsymbol{B}(0.4^{|i-j|^{1/3}})\boldsymbol{B}$, where

$$\boldsymbol{B} = \text{diag}[\{0.5 + 1/(p+1)\}^{1/2}, ..., \{0.5 + p/(p+1)\}^{1/2}]. \qquad (3)$$

Note that $\text{tr}(\boldsymbol{\Sigma}_i) = p$ ($i = 1, 2$). Also, note that $\boldsymbol{\Sigma}_i$, $i = 1, 2$, hold (A-vi). We considered two cases: (i) $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$, and (ii) $\boldsymbol{\mu}_1 = ((3/p)^{1/2}, ..., (3/p)^{1/2})^T$ and $\boldsymbol{\mu}_2 = \boldsymbol{0}$ (i.e., $||\boldsymbol{\mu}||^2 = ||\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2||^2 = 3$). Let $\widetilde{T} = (\widehat{T}_{\boldsymbol{n}} - ||\boldsymbol{\mu}||^2)/\widehat{K}^{1/2}$. Note that $\widetilde{T} = (||\boldsymbol{T_n} - \boldsymbol{\mu}||^2 - \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{n}})/\widehat{K}^{1/2}$ for case (i). We considered the cases of $p = 10$, 100 and 1000. We generated $\boldsymbol{x}_{ij} - \boldsymbol{\mu}_i$, $j = 1, 2, ...,$ ($i = 1, 2$) independently from a p-variate distribution. We considered three distributions: (a) $N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$, (b) p-variate t-distribution, $t_p(\boldsymbol{0}, \boldsymbol{\Sigma}_i, \nu)$ with mean zero, covariance matrix $\boldsymbol{\Sigma}_i$ and degrees of freedom $\nu = 10$, and (c) $\boldsymbol{\Gamma}_i \boldsymbol{w}_{ij}$, where $\boldsymbol{w}_{ij} = (w_{i1j}, ..., w_{ir_ij})^T$ and $w_{ii'j} = (y_{ii'j} - 1)/\sqrt{2}$ ($i' = 1, ..., r_i$) in which $y_{ii'j}$s are i.i.d. as the chi-squared distribution with 1 degree of freedom. Here, $r_i = 2p$ and $\boldsymbol{\Gamma}_i$

is a $p \times r_i$ matrix such that $\mathbf{\Gamma}_i = [\mathbf{H}_i \mathbf{\Lambda}_i^{1/2}/\sqrt{2}, -\mathbf{H}_i \mathbf{\Lambda}_i^{1/2}/\sqrt{2}]$. Note that $\mathbf{\Gamma}_i \mathbf{\Gamma}_i^T = \mathbf{\Sigma}_i$. It should be noted that (A-v) or (A-v') is met in (b) when ν is sufficiently large. Also, (A-v) is met in (c).

Independent pseudorandom 2000 observations of $\widetilde{T}$ were generated from each distribution. Let $\widetilde{T}_r$ be the rth observation of $\widetilde{T}$ for $r = 1, ..., 2000$. In the end of the rth replication, we checked whether the inequality, $\widetilde{T}_r \leq z_{0.05}$, is true (or false) and defined $P_{ir} = 1$ (or 0), where $z_{0.05}$ is the upper 0.05 point of $N(0, 1)$. We calculated $\overline{P}(0.95) = \sum_{r=1}^{2000} P_r/2000$ as an estimate of $P_\theta(\widetilde{T} \leq z_{0.05})$. Note that the standard deviation of the estimates is less than 0.011. From Corollaries 4 and 5, it holds that $P_\theta(\widetilde{T} \leq z_{0.05}) \to 0.95$. In Figs. 1 and 2, we gave the histograms of $\widetilde{T}$ together with $\overline{P}(0.95)$ for (a), (b) and (c) when $p = 10, 100$ and 1000. We considered case (i) in Fig. 1 and case (ii) in Fig. 2. From Corollaries 4 and 5, we also displayed the
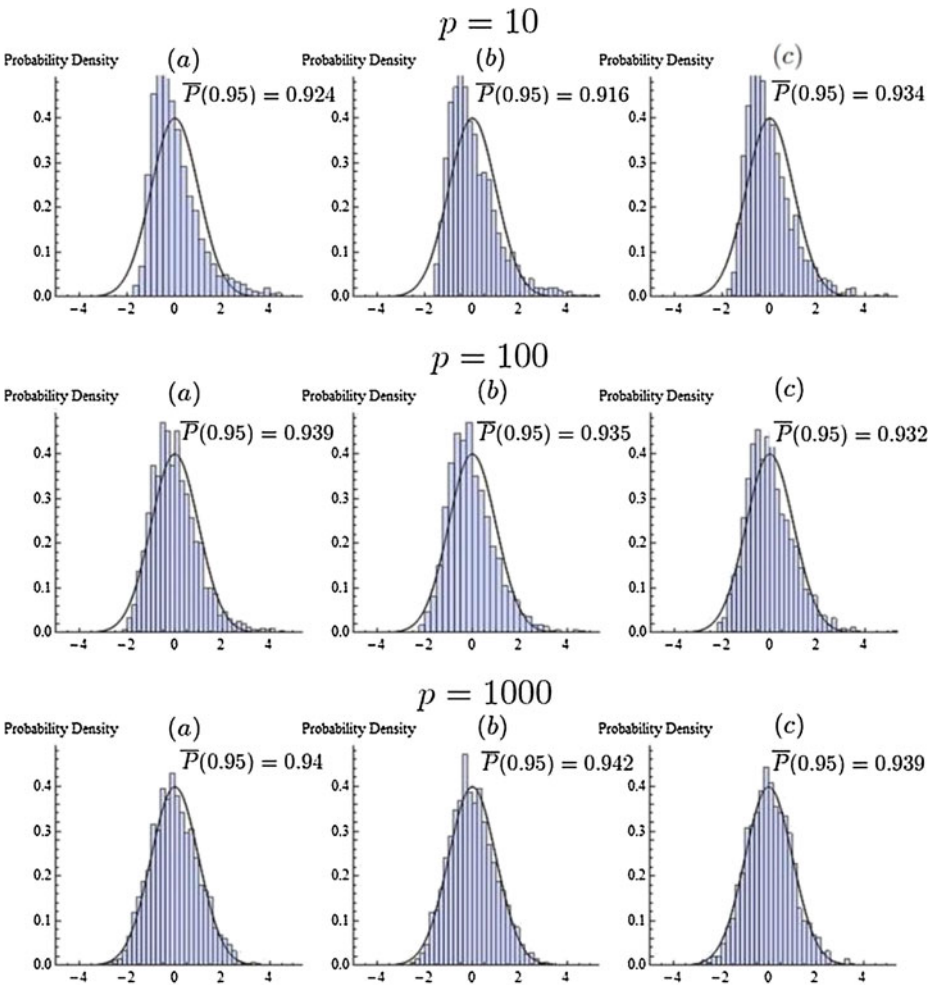
Fig. 1 The histograms of $\widetilde{T}$ together with $\overline{P}(0.95)$ and the probability density of $N(0, 1)$ for **(a)**, **(b)** and **(c)** when $p = 10, 100$ and 1000 in case of (i)
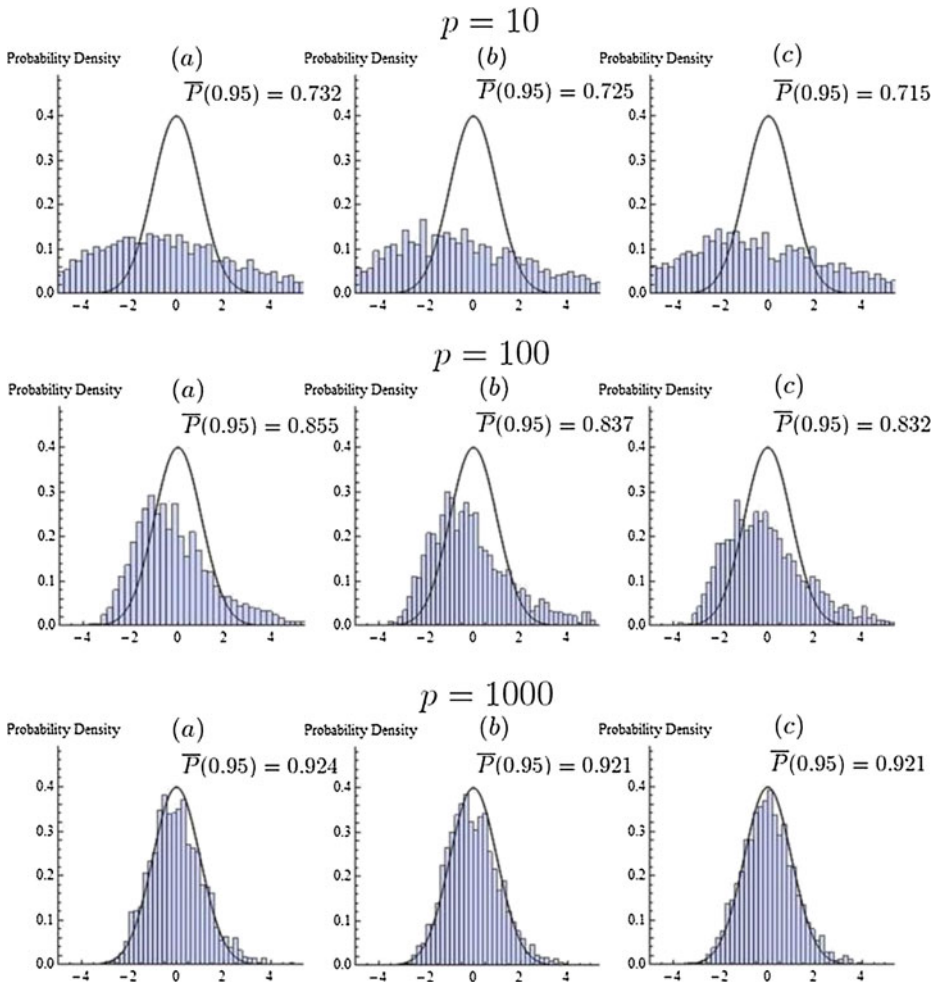
Fig. 2 The histograms of $\widetilde{T}$ together with $\overline{P}(0.95)$ and the probability density of $N(0, 1)$ for (**a**), (**b**) and (**c**) when $p =$10, 100 and 1000 in case of (ii)

asymptotic probability density of $\widetilde{T}$, $N(0, 1)$. We observed from Fig. 1 that the histograms become close to the probability density and $\overline{P}(0.95)$s become close to 0.95 as p increases in case (i). Compared to (i), in view of Fig. 2, the convergence of those quantities seemed to be slow in case (ii) when p is not large enough to meet $K_*/K \to 1$. However, for sufficiently large p, we observed that they give adequate performances even in non-Gaussian cases such as (b) and (c).

3 Estimation of tr(Σ^2)

Throughout this section, we omit the subscript with regard to the population. Yata and Aoshima (2013) developed the *extended cross-data-matrix (ECDM)*

methodology that is an extension of the CDM methodology created by Yata and Aoshima ([2010](#)). The ECDM methodology can be applied to obtaining an unbiased estimator of $\text{tr}(\boldsymbol{\Sigma}^2)$ as follows: We assume $n \geq 4$. Let $n_{(1)} = \lceil n/2 \rceil$ and $n_{(2)} = n - n_{(1)}$, where $\lceil x \rceil$ denotes the smallest integer $\geq x$. Let

$$
V_{n(1)(k)} = \begin{cases} \{\lfloor k/2 \rfloor - n_{(1)} + 1, ..., \lfloor k/2 \rfloor\} & \text{if } \lfloor k/2 \rfloor \geq n_{(1)}, \\ \{1, ..., \lfloor k/2 \rfloor\} \cup \{\lfloor k/2 \rfloor + n_{(2)} + 1, ..., n\} & \text{otherwise;} \end{cases}
$$

$$
V_{n(2)(k)} = \begin{cases} \{\lfloor k/2 \rfloor + 1, ..., \lfloor k/2 \rfloor + n_{(2)}\} & \text{if } \lfloor k/2 \rfloor \leq n_{(1)}, \\ \{1, ..., \lfloor k/2 \rfloor - n_{(1)}\} \cup \{\lfloor k/2 \rfloor + 1, ..., n\} & \text{otherwise} \end{cases}
$$

for $k = 3, ..., 2n - 1$, where $\lfloor x \rfloor$ denotes the largest integer $\leq x$. Let $\#(\boldsymbol{S})$ denote the number of elements in a set $\boldsymbol{S}$. Note that $\#(\boldsymbol{V}_{n(l)(k)}) = n_{(l)}$, $l = 1, 2$, $\boldsymbol{V}_{n(1)(k)} \cap \boldsymbol{V}_{n(2)(k)} = \emptyset$ and $\boldsymbol{V}_{n(1)(k)} \cup \boldsymbol{V}_{n(2)(k)} = \{1, ..., n\}$ for $k = 3, ..., 2n - 1$. Also, note that $i \in \boldsymbol{V}_{n(1)(i+j)}$ and $j \in \boldsymbol{V}_{n(2)(i+j)}$ for $i < j (\leq n)$. Let

$$
\overline{\boldsymbol{x}}_{n(1)(k)} = n_{(1)}^{-1} \sum_{j \in \boldsymbol{V}_{n(1)(k)}} \boldsymbol{x}_j \quad \text{and} \quad \overline{\boldsymbol{x}}_{n(2)(k)} = n_{(2)}^{-1} \sum_{j \in \boldsymbol{V}_{n(2)(k)}} \boldsymbol{x}_j
$$

for $k = 3, ..., 2n - 1$. Then, Yata and Aoshima ([2013](#)) gave an estimator of $\text{tr}(\boldsymbol{\Sigma}^2)$ by

$$
W_n = \frac{2u_n}{n(n-1)} \sum_{i<j}^{n} \left\{ (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{n(1)(i+j)})^T (\boldsymbol{x}_j - \overline{\boldsymbol{x}}_{n(2)(i+j)}) \right\}^2, \tag{4}
$$

where $u_n = n_{(1)} n_{(2)} / \{(n_{(1)} - 1)(n_{(2)} - 1)\}$. Note that $E_{\boldsymbol{\theta}}(W_n) = \text{tr}(\boldsymbol{\Sigma}^2)$. Aoshima and Yata (A distance-based, misclassification rate adjusted classifier for multiclass, high-dimensional data. To appear in Ann Inst Stat Math) and Yata and Aoshima ([2013](#)) showed that

$$
\text{Var}_{\boldsymbol{\theta}} \left(\frac{W_n}{\text{tr}(\boldsymbol{\Sigma}^2)} \right) = \frac{4}{n^2} \{1 + o(1)\} + O \left\{ \frac{\text{tr}(\boldsymbol{\Sigma}^4)}{\text{tr}(\boldsymbol{\Sigma}^2)^2 n} \right\} \to 0 \tag{5}
$$

as $p \to \infty$ and $n \to \infty$ under (A-v). On the other hand, under (A-v'), we can claim as $p \to \infty$ and $n \to \infty$ that

$$
\text{Var}_{\boldsymbol{\theta}} \left(\frac{W_n}{\text{tr}(\boldsymbol{\Sigma}^2)} \right) = O(n^{-2}) + O \left\{ \frac{\text{tr}(\boldsymbol{\Sigma}^4)}{\text{tr}(\boldsymbol{\Sigma}^2)^2 n} \right\} \to 0.
$$

Remark 3 Assume (A-i). It holds as $p \to \infty$ and $n \to \infty$ that

$$
\text{Var}_{\boldsymbol{\theta}} \left(\frac{W_n}{\text{tr}(\boldsymbol{\Sigma}^2)} \right) = \frac{4}{n^2} \{1 + o(1)\} + 8 \frac{\text{tr}(\boldsymbol{\Sigma}^4)}{\text{tr}(\boldsymbol{\Sigma}^2)^2 n} \{1 + o(1)\}. \tag{6}
$$

On the other hand, Bai and Sarandasa ([1996](#)) and Srivastava ([2005](#)) considered an estimator of $\text{tr}(\boldsymbol{\Sigma}^2)$ by $V_n = c_n^{-1} \{\text{tr}(\boldsymbol{S}_n^2) - \text{tr}(\boldsymbol{S}_n)^2 / (n-1)\}$ with $c_n = (n-2)(n+1)/(n-1)^2$. They showed that, when the population distribution is Gaussian, it holds that $E_{\boldsymbol{\theta}}(V_n) = \text{tr}(\boldsymbol{\Sigma}^2)$ together with Eq. [6](#). It should be noted that V_n is not an unbiased estimator unless the population distribution is Gaussian. In addition, one cannot claim $\text{Var}_{\boldsymbol{\theta}} \{V_n / \text{tr}(\boldsymbol{\Sigma}^2)\} < \infty$ unless the eighth moments of each variable in $\boldsymbol{z}_j$ are uniformly bounded.

4 Applications

In this section, we give several applications of the results in Section 2.

4.1 Confidence Regions for μ

Let z_α be a constant such that $P\{N(0, 1) > z_\alpha\} = \alpha$. We construct confidence regions for μ by

$$\boldsymbol{R} = \left\{ \boldsymbol{\mu} \in R^p \mid \max\left\{ \widehat{\Sigma}_{\mathbf{n}} - z_{\alpha/2}\widehat{K}^{1/2}, \ 0 \right\} \leq ||\boldsymbol{T}_{\mathbf{n}} - \boldsymbol{\mu}||^2 \leq \widehat{\Sigma}_{\mathbf{n}} + z_{\alpha/2}\widehat{K}^{1/2} \right\};$$

$$\boldsymbol{R}_U = \left\{ \boldsymbol{\mu} \in R^p \mid ||\boldsymbol{T}_{\mathbf{n}} - \boldsymbol{\mu}||^2 \leq \widehat{\Sigma}_{\mathbf{n}} + z_\alpha\widehat{K}^{1/2} \right\},$$

where $\alpha \in (0, 1/2)$. Then, from Corollary 4, it holds as $p \to \infty$ and $n_i \to \infty$, $i = 1, ..., k$, that

$$P_{\boldsymbol{\theta}}(\boldsymbol{\mu} \in \boldsymbol{R}) = 1 - \alpha + o(1) \quad \text{and} \quad P_{\boldsymbol{\theta}}(\boldsymbol{\mu} \in \boldsymbol{R}_U) = 1 - \alpha + o(1)$$

under (A-v) and (A-vi).

Remark 4 When $\widehat{\Sigma}_{\mathbf{n}} > z_{\alpha/2}\widehat{K}^{1/2}$, $\boldsymbol{R}$ indicates that $\boldsymbol{\mu}$ is included in the region sandwiched by the two p-dimensional spheres with radii of $(\widehat{\Sigma}_{\mathbf{n}} - z_{\alpha/2}\widehat{K}^{1/2})^{1/2}$ and $(\widehat{\Sigma}_{\mathbf{n}} + z_{\alpha/2}\widehat{K}^{1/2})^{1/2}$ from center $\boldsymbol{T}_{\mathbf{n}}$. See Section 2 in Aoshima and Yata (2011a) for the details.

4.2 Confidence Intervals for $||\boldsymbol{\mu}||^2$

We construct confidence intervals for $||\boldsymbol{\mu}||^2$ by

$$I = \left[\max\left\{ \widehat{T}_{\mathbf{n}} - z_{\alpha/2}\widehat{K}^{1/2}, \ 0 \right\}, \ \widehat{T}_{\mathbf{n}} + z_{\alpha/2}\widehat{K}^{1/2} \right];$$

$$I_L = \left[\max\left\{ \widehat{T}_{\mathbf{n}} - z_\alpha\widehat{K}^{1/2}, \ 0 \right\}, \ \infty \right),$$

where $\alpha \in (0, 1/2)$. Then, from Corollary 5, it holds as $p \to \infty$ and $n_i \to \infty$, $i = 1, ..., k$, that

$$P_{\boldsymbol{\theta}}(||\boldsymbol{\mu}||^2 \in I) = 1 - \alpha + o(1) \quad \text{and} \quad P_{\boldsymbol{\theta}}(||\boldsymbol{\mu}||^2 \in I_L) = 1 - \alpha + o(1)$$

under (A-v) to (A-vii). We emphasize that one can apply I and I_L to the discriminant analysis for high-dimensional data. Refer to Section 4 in Aoshima and Yata (2011a, c) and Sections 3 and 4 in Aoshima and Yata (A distance-based, misclassification rate adjusted classifier for multiclass, high-dimensional data. To appear in Ann Inst Stat Math).

4.3 Test of $\mu = 0$ Against $\mu \neq 0$

We consider the following test:

$$H_0 : \boldsymbol{\mu} = \boldsymbol{0} \quad \text{vs.} \quad H_1 : \boldsymbol{\mu} \neq \boldsymbol{0}. \tag{7}$$

For given $\alpha \in (0, 1/2)$, we test the hypothesis (7) by

$$\text{rejecting } H_0 \Longleftrightarrow \frac{\widehat{T}_{\mathbf{n}}}{\widehat{K}^{1/2}} > z_\alpha.$$

Then, from Corollary 5, it holds as $p \to \infty$ and $n_i \to \infty$, $i = 1, ..., k$, that

$$size = \alpha + o(1) \quad \text{and} \quad power = \Phi\Big(\frac{||\boldsymbol{\mu}||^2}{K^{1/2}} - z_\alpha\Big) + o(1) \tag{8}$$

under (A-v) to (A-vii), where $\Phi(\cdot)$ denotes the c.d.f. of $N(0, 1)$. If one cannot assume (A-vii) in case of $\boldsymbol{\mu} \neq \mathbf{0}$, from Corollaries 2 and 5, it holds as $p \to \infty$ and $n_i \to \infty$, $i = 1, ..., k$, that

$$size = \alpha + o(1) \quad \text{and} \quad power = 1 + o(1)$$

under (A-v), (A-vi) and (A-viii).

Remark 5 When $k = 2$ and $(b_1, b_2) = (1, -1)$, Chen and Qin (2010) also gave Eq. 8 under slightly different conditions.

Remark 6 After replacing (A-v) with (A-v'), all the results in Section 4 are still justified.

5 Sample Size Determination

In this section, we provide sample size determination so as to satisfy prespecified accuracy on inference.

5.1 Given-Bandwidth Confidence Region for $\boldsymbol{\mu}$

We consider constructing a given-bandwidth confidence region for $\boldsymbol{\mu}$ by

$$\boldsymbol{R_n} = \big\{\boldsymbol{\mu} \in R^p \mid \max\{\widehat{\Sigma}_\mathbf{n} - \delta, \ 0\} \le ||\boldsymbol{T_n} - \boldsymbol{\mu}||^2 \le \widehat{\Sigma}_\mathbf{n} + \delta\big\} \tag{9}$$

for given δ (> 0). We assume $\delta = o\{\min_{i=1,...,k} \text{tr}(\Sigma_i^2)^{1/2}\}$. For given $\alpha \in (0, 1/2)$, we determine the sample size so as to satisfy

$$P_\theta(\boldsymbol{\mu} \in \boldsymbol{R_n}) \ge 1 - \alpha.$$

Aoshima and Yata (2011a) considered sample size determination as follows: From the fact that $\text{tr}(\Sigma_i \Sigma_j) \le \{\text{tr}(\Sigma_i^2)\text{tr}(\Sigma_j^2)\}^{1/2}$, it holds $K^{1/2} \le \sqrt{2} \sum_{i=1}^k b_i^2 \ \text{tr}(\Sigma_i^2)^{1/2}/ (n_i - 1)$. One may choose n_is such that

$$\min \sum_{i=1}^k n_i \quad \text{subject to} \quad \sqrt{2} \sum_{i=1}^k b_i^2 \text{tr}(\Sigma_i^2)^{1/2}/(n_i - 1) \le \delta/z_{\alpha/2}.$$

Then, the sample size is determined by

$$n_i \ge \frac{z_{\alpha/2}\sqrt{2}}{\delta} |b_i| \text{tr}(\Sigma_i^2)^{1/4} \sum_{j=1}^k |b_j| \text{tr}(\Sigma_j^2)^{1/4} + 1 \quad (= C_i, \text{ say}) \tag{10}$$

for each π_i. Note that $n_i \to \infty$, $i = 1, ..., k$, as $p \to \infty$.

Theorem 6 (Aoshima and Yata 2011a) *Assume* (A-iv) *and either* (A-ii) *or* (A-iii). *Then, for n_is satisfying Eq.* 10, *it holds as $p \to \infty$ that*

$$\liminf P_\theta(\mu \in R_n) \geq 1 - \alpha. \tag{11}$$

We can claim Eq. 11 under mild conditions as follows.

Theorem 7 *Assume* (A-v) *and* (A-vi). *Then, for n_is satisfying Eq.* 10, *Eq.* 11 *holds as $p \to \infty$.*

5.2 Two-Stage Procedure

Since Σ_is are unknown, it is necessary to estimate C_i in Eq. 10. We proceed with the following two steps along the lines of the elastic two-stage procedure given by Aoshima and Yata (2011b):

1. Choose $m_i(\geq 4)$, $i = 1, ..., k$, such as

$$\frac{m_i}{C_i} \leq 1, \quad \frac{C_i}{m_i^2} \to 0 \text{ and } \frac{C_i}{m_i}\frac{\text{tr}(\Sigma_i^4)}{\text{tr}(\Sigma_i^2)^2} \to 0 \text{ as } p \to \infty \text{ under (A-vi)}. \tag{12}$$

   Note that Eq. 12 is met when $\liminf_{p\to\infty} m_i/C_i > 0$ and $\limsup_{p\to\infty} m_i/C_i < 1$. Take pilot samples, x_{ij}, $j = 1, ..., m_i$, of size m_i from each π_i. Then, calculate W_{im_i} according to Eq. 4. Define the total sample size for each π_i by

$$N_i = \max\left\{ m_i, \left\lceil \frac{z_{\alpha/2}\sqrt{2}}{\delta}|b_i|W_{im_i}^{1/4}\sum_{j=1}^{k}|b_j|W_{jm_j}^{1/4} \right\rceil + 1 \right\}, \tag{13}$$

   where $\lceil x \rceil$ denotes the smallest integer $\geq x$.
2. For each i, if $N_i = m_i$, do not take any additional samples from π_i and otherwise, that is if $N_i > m_i$, take additional samples, x_{ij}, $j = m + 1, ..., N_i$, of size $N_i - m_i$ from π_i. By combining the initial samples and the additional samples, calculate $\bar{x}_{iN_i}$ and S_{iN_i}, $i = 1, ..., k$. Let $N = (N_1, ..., N_k)$. Then, define R_N according to Eq. 9 with $T_N = \sum_{i=1}^{k} b_i\bar{x}_{iN_i}$ and $\widehat{\Sigma}_N = \sum_{i=1}^{k} b_i^2\text{tr}(S_{iN_i})/N_i$.

We have the following theorem.

Theorem 8 *Assume* (A-v) *and* (A-vi). *For two-stage procedure given by Eqs.* 12 *and* 13, *we have as $p \to \infty$ that*

$$\liminf P_\theta(\mu \in R_N) \geq 1 - \alpha.$$

Remark 7 The results in Theorems 7 and 8 are still justified under (A-v') instead of (A-v).

Remark 8 Under Eq. 12, (A-v) and (A-vi), it holds as $p \to \infty$ that $\text{Var}_\theta\{W_{im_i}/\text{tr}(\Sigma_i^2)\} = o(C_i^{-1})$. Then, we can claim as $p \to \infty$ that $N_i/C_i = 1 + o_p(1)$, which is in the HDLSS situation in the sense that $N_i/p = o_p(1)$ under the condition that $\max_{i=1,...,k} \text{tr}(\Sigma_i^2)^{1/2}/\delta = o(p)$.

Remark 9 One may choose $m_i (\geq 4)$ such as $m_i/C_i > 1$ for some i. Then, the assertion in Theorem 8 can still be claimed. However, it may cause over-sampling in the sense that $N_i/C_i > 1$ w.p.1.

Remark 10 One can obtain sample size determination both for the confidence interval (Section 4.2) and for the test (Section 4.3) as well in similar fashion. Under (A-v) and (A-vi), it can be verified to ensure the accuracy required in Yata and Aoshima (2012, Sections 2 and 3) and Aoshima and Yata (2011a, Section 3). We omit the details for brevity.

5.3 Simulation

In order to study the performance of the two-stage procedure given by Eqs. 12 and 13, we used computer simulations. Our goal was to construct a 95 % given-bandwidth confidence region, $\boldsymbol{R_N}$. In other words, we set $\alpha = 0.05$. We set $\boldsymbol{\mu}_{i'} = (0, ..., 0)^T$, $b_{i'} = (-1)^{i'-1}$ and $\boldsymbol{\Sigma}_{i'} = c_{i'} \boldsymbol{B}(\rho_{i'}^{|i-j|^{1/3}}) \boldsymbol{B}$, $i' = 1, ..., k$, where $\boldsymbol{B}$ is defined by Eq. 3. Note that $\mathrm{tr}(\boldsymbol{\Sigma}_i) = c_i p$ $(i = 1, .., k)$. We considered two cases: (i) $k = 2$, $\delta = 5$, $(c_1, c_2) = (1, 1)$, $(\rho_1, \rho_2) = (0.3, 0.4)$ and $(m_1, m_2) = (10, 10)$, and (ii) $k = 4$, $\delta = 10$, $(c_1, c_2, c_3, c_4) = (1, 1, 1.2, 1.2)$, $(\rho_1, \rho_2, \rho_3, \rho_4) = (0.3, 0.4, 0.3, 0.4)$ and $(m_1, m_2, m_3, m_4) = (10, 10, 15, 15)$. In Table 1, we generated independent pseudorandom observations from $\pi_i : N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$, $i = 1, ..., k$. In Table 2, we generated them from $\boldsymbol{\Gamma}_i \boldsymbol{w}_{ij}$, where $\boldsymbol{\Gamma}_i = \boldsymbol{H}_i \boldsymbol{\Lambda}_i^{1/2}$, $\boldsymbol{w}_{ij} = (w_{i1j}, ..., w_{ipj})^T$ and $w_{i i' j} = y_{i i' j} - 1$ $(i' = 1, ..., p)$ in which $y_{i i' j}$s are i.i.d. as the Poisson distribution with parameter $\lambda = 1$ for π_i, $i = 1, ..., k$.

Table 1 Required sample size and coverage probability given by Eqs. 12 and 13 when $\pi_i : N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$, $i = 1, ..., k$

	C_i	$\bar{n}_i$	$\bar{n}_i - C_i$	$V(n_i)$	$\overline{P}$	$s(\overline{P})$
Case (i): $k = 2$, $\delta = 5$ and $(m_1, m_2) = (10, 10)$						
$p = 200$						
π_1	22.67	22.67	0.0	8.42	0.957	0.00456
π_2	25.29	25.25	−0.04	17.51		
$p = 1000$						
π_1	49.87	49.55	−0.31	30.86	0.954	0.00468
π_2	56.1	55.82	−0.28	47.26		
Case (ii): $k = 4$, $\delta = 10$ and $(m_1, m_2, m_3, m_4) = (10, 10, 15, 15)$						
$p = 200$						
π_1	23.7	23.76	0.06	6.29	0.958	0.00449
π_2	26.45	26.26	−0.19	11.32		
π_3	25.87	25.96	0.09	3.89		
π_4	28.88	28.88	0.0	8.82		
$p = 1000$						
π_1	52.2	51.97	−0.23	21.6	0.95	0.0049
π_2	58.73	58.26	−0.47	33.03		
π_3	57.09	57.07	−0.02	12.55		
π_4	64.24	63.96	−0.28	21.95		

Table 2 Required sample size and coverage probability given by Eqs. 12 and 13 when $\boldsymbol{\Gamma}_i = \boldsymbol{H}_i \boldsymbol{\Lambda}_i^{1/2}$, $\boldsymbol{w}_{ij} = (w_{i1j}, ..., w_{ipj})^T$ and $w_{ii'j} = y_{ii'j} - 1$ $(i' = 1, ..., p)$ in which $y_{ii'j}$s are i.i.d. as the Poisson distribution with parameter $\lambda = 1$ for π_i, $i = 1, ..., k$

	C_i	$\overline{n}_i$	$\overline{n}_i - C_i$	$V(n_i)$	$\overline{P}$	$s(\overline{P})$
Case (i): $k = 2$, $\delta = 5$ and $(m_1, m_2) = (10, 10)$						
$p = 200$						
π_1	22.67	22.51	−0.16	10.06	0.958	0.00451
π_2	25.29	25.06	−0.23	19.89		
$p = 1000$						
π_1	49.87	49.74	−0.13	32.03	0.955	0.00466
π_2	56.1	55.89	−0.21	50.49		
Case (ii): $k = 4$, $\delta = 10$ and $(m_1, m_2, m_3, m_4) = (10, 10, 15, 15)$						
$p = 200$						
π_1	23.7	23.65	−0.05	6.65	0.954	0.00471
π_2	26.45	26.15	−0.3	14.02		
π_3	25.87	25.97	0.1	5.06		
π_4	28.88	28.84	−0.04	11.78		
$p = 1000$						
π_1	52.2	52.02	−0.18	22.79	0.955	0.00466
π_2	58.73	58.52	−0.21	40.09		
π_3	57.09	57.11	−0.02	13.63		
π_4	64.24	64.16	−0.08	24.75		

When $p = 200$ and 1000, we used the two-stage procedure given by Eqs. 12 and 13. The findings were obtained by averaging the outcomes from 2000 ($= R$, say) replications. Under a fixed scenario, suppose that the rth replication ends with $N_i = n_{ir}$ $(i = 1, ..., k)$ observations and the corresponding confidence region with $\boldsymbol{n}_r = (n_{1r}, ..., n_{kr})$ for $r = 1, ..., R$. Let $\overline{n}_i = R^{-1} \sum_{r=1}^{R} n_{ir}$ and $V(n_i) = (R-1)^{-1} \sum_{r=1}^{R} (n_{ir} - \overline{n}_i)^2$. In the end of the rth replication, we checked whether $\boldsymbol{\mu}$ does (or does not) belong to the corresponding confidence region and defined $P_r = 1$ (or 0) accordingly. Let $\overline{P} = R^{-1} \sum_{r=1}^{R} P_r$, which estimates the target coverage probability, having its estimated standard error $s(\overline{P})$ where $s^2(\overline{P}) = R^{-1} \overline{P}(1 - \overline{P})$.

Let us explain, for example, the entries from the second block for case (i) in Table 1 that were given when $p = 1000$. We had $C_1 = 49.87$ and $C_2 = 56.1$ from Eq. 10. From 2000 independent replications, we observed $\overline{n}_1 = 49.55$ ($\overline{n}_1 - C_1 = -0.31$), $\overline{n}_2 = 55.82$ ($\overline{n}_2 - C_2 = -0.28$) and $\overline{P} = 0.954$ together with $V(n_1) = 30.86$, $V(n_2) = 47.26$ and $s(\overline{P}) = 0.00468$. Throughout, the two-stage procedure seemed to construct required confidence regions successfully even for a discrete case such as in Table 2.

6 Example

We analyzed gene expression data given by Chiaretti et al. (2004) in which the data set consists of 12625 ($= p$) genes. The expression measures were obtained by using the three-step robust multichip average (RMA) preprocessing method. Refer to Pollard et al. (2005) as well for the details. The data set had two tumor cellular subtypes, B-cell and T-cell. We divided each type into two groups with

respect to the relapse as follows: (a) π_1: B-cell with the relapse ($n_1 = 50$ samples); (b) π_2: B-cell without the relapse ($n_2 = 26$ samples); (c) π_3: T-cell with the relapse ($n_3 = 15$ samples); and (d) π_4: T-cell without the relapse ($n_4 = 9$ samples).

6.1 Tests of $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_3$ and $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_4$

We first considered testing $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_3$. We set $\alpha = 0.05$. We calculated $\widehat{T}_{\mathbf{n}} = ||\bar{\boldsymbol{x}}_{1n_1} - \bar{\boldsymbol{x}}_{3n_3}||^2 - \mathrm{tr}(\boldsymbol{S}_{1n_1})/n_1 - \mathrm{tr}(\boldsymbol{S}_{3n_3})/n_3 = 1267$ and $\widehat{K} = 2(W_{1n_1}/\{n_1(n_1 - 1)\} + W_{3n_3}/\{n_3(n_3 - 1)\}) + 4\mathrm{tr}(\boldsymbol{S}_{1n_1}\boldsymbol{S}_{3n_3})/(n_1 n_3) = 4331$ with $W_{1n_1} = 2.84 \times 10^5$, $W_{3n_3} = 2.83 \times 10^5$ and $\mathrm{tr}(\boldsymbol{S}_{1n_1}\boldsymbol{S}_{3n_3}) = 2.64 \times 10^5$ according to Eq. 4. Hence, we obtained $\widehat{T}_{\mathbf{n}}/\widehat{K}^{1/2} = 19.3$. From the fact that

$$\frac{\widehat{T}_{\mathbf{n}}}{\widehat{K}^{1/2}} > z_\alpha = 1.64,$$

we rejected the null hypothesis, $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_3$, with size 0.05 according to Section 4.3. According to Section 4.2, we calculated 95 % confidence intervals for $||\boldsymbol{\mu}||^2 = ||\boldsymbol{\mu}_1 - \boldsymbol{\mu}_3||^2$ as follows:

$$I = [1138, 1396],$$
$$I_L = \big[\max\big\{\widehat{T}_{\mathbf{n}} - z_\alpha \widehat{K}^{1/2}, \, 0\big\}, \infty\big) = [1159, \infty).$$

Since both I and I_L did not include $||\boldsymbol{\mu}|| = 0$, we concluded $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_3$ significantly.
Similarly, we considered testing $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_4$. From the fact that

$$\frac{\widehat{T}_{\mathbf{n}}}{\widehat{K}^{1/2}} = \frac{||\bar{\boldsymbol{x}}_{2n_2} - \bar{\boldsymbol{x}}_{4n_4}||^2 - \mathrm{tr}(\boldsymbol{S}_{2n_2})/n_2 - \mathrm{tr}(\boldsymbol{S}_{4n_4})/n_4}{\sqrt{2W_{2n_2}/\{n_2(n_2 - 1)\} + 2W_{4n_4}/\{n_4(n_4 - 1)\} + 4\mathrm{tr}(\boldsymbol{S}_{2n_2}\boldsymbol{S}_{4n_4})/(n_2 n_4)}}$$
$$= 23.6 > z_\alpha,$$

we rejected the null hypothesis, $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_4$, with size $\alpha = 0.05$.

6.2 Confidence Region for $\boldsymbol{\mu} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_3) - (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_4)$

We considered constructing a confidence region for $\boldsymbol{\mu} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_3) - (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_4)$. We set $\alpha = 0.05$. We calculated $\boldsymbol{T}_{\mathbf{n}} = (\bar{\boldsymbol{x}}_{1n_1} - \bar{\boldsymbol{x}}_{3n_3}) - (\bar{\boldsymbol{x}}_{2n_2} - \bar{\boldsymbol{x}}_{4n_4}) = (0.151, -0.193, ..., 0.115, -0.102)^T$, $\widehat{K} = 2\sum_{i=1}^{4} W_{in_i}/\{n_i(n_i - 1)\} + 4\sum_{i<j} \mathrm{tr}(\boldsymbol{S}_{in_i}\boldsymbol{S}_{jn_j})/(n_i n_j) = 20010$ and $\widehat{\Sigma}_{\mathbf{n}} = \sum_{i=1}^{4} \mathrm{tr}(\boldsymbol{S}_{in_i})/n_i = 563$. According to Section 4.1, we obtained a confidence region as

$$\boldsymbol{R} = \big\{\boldsymbol{\mu} \in R^p \mid \max\big\{\widehat{\Sigma}_{\mathbf{n}} - z_{\alpha/2}\widehat{K}^{1/2}, \, 0\big\} \leq ||\boldsymbol{T}_{\mathbf{n}} - \boldsymbol{\mu}||^2 \leq \widehat{\Sigma}_{\mathbf{n}} + z_{\alpha/2}\widehat{K}^{1/2}\big\}$$
$$= \big\{\boldsymbol{\mu} \in R^p \mid 285 \leq ||\boldsymbol{T}_{\mathbf{n}} - \boldsymbol{\mu}||^2 \leq 840\big\}.$$

When $\boldsymbol{\mu} = \boldsymbol{0}$, we had $||\boldsymbol{T}_{\mathbf{n}} - \boldsymbol{\mu}||^2 = ||\boldsymbol{T}_{\mathbf{n}}||^2 = 651$, so that $\boldsymbol{\mu} = \boldsymbol{0} \in \boldsymbol{R}$. Thus we considered $\boldsymbol{0}$ as a likely candidate of $\boldsymbol{\mu} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_3) - (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_4)$. In other words, we

concluded that there is not much significant difference between B-cell and T-cell in terms of the relapse.

6.3 Given-Bandwidth Confidence Region for $\boldsymbol{\mu} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$

We considered constructing a given-bandwidth confidence region for $\boldsymbol{\mu} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ along the line of Section 5.2. We set $\alpha = 0.05$, $\delta = 150$ and $m_1 = m_2 = 15$. By using pilot samples of sizes $m_1 = m_2 = 15$, we calculated $W_{1m_1} = 2.96 \times 10^5$ and $W_{2m_2} = 2.21 \times 10^5$ according to Eq. 4. From Eq. 13, we calculated the total sample sizes as

$$N_1 = \max \left\{ 15, \ \left\lceil \left(\frac{z_{\alpha/2}\sqrt{2}}{\delta} W_{1m_1}^{1/4} \sum_{j=1}^{2} W_{jm_j}^{1/4} \right)^2 \right\rceil + 1 \right\} = 21 \quad \text{and} \quad N_2 = 20.$$

So, we took the next 6 samples from π_1 and the next 5 samples from π_2. Then, we had $\boldsymbol{T_N} = \bar{\boldsymbol{x}}_{1N_1} - \bar{\boldsymbol{x}}_{2N_2} = (0.101, -0.185, ..., 0.012, 0.025)^T$ and $\widehat{\Sigma}_{\mathbf{N}} = \text{tr}(\boldsymbol{S}_{1N_1})/N_1 + \text{tr}(\boldsymbol{S}_{2N_2})/N_2 = 259$. According to Eq. 9, we constructed a given-bandwidth confidence region as

$$\boldsymbol{R_N} = \left\{ \boldsymbol{\mu} \in R^p \mid \max\{\widehat{\Sigma}_{\mathbf{N}} - \delta, \ 0\} \le ||\boldsymbol{T_N} - \boldsymbol{\mu}||^2 \le \widehat{\Sigma}_{\mathbf{N}} + \delta \right\}$$
$$= \left\{ \boldsymbol{\mu} \in R^p \mid 109 \le ||\boldsymbol{T_N} - \boldsymbol{\mu}||^2 \le 409 \right\}.$$

When $\boldsymbol{\mu} = \boldsymbol{0}$, we had $||\boldsymbol{T_N} - \boldsymbol{\mu}||^2 = ||\boldsymbol{T_N}||^2 = 243$, so that $\boldsymbol{\mu} = \boldsymbol{0} \in \boldsymbol{R_N}$. Thus we considered $\boldsymbol{0}$ as a likely candidate of $\boldsymbol{\mu} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. If one considers a mean structure such as $\boldsymbol{\mu} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = c(1, ..., 1)^T$ with an unknown constant c, we can provide an allowable range of c by $c \in [-0.126, 0.104]$ such that $\boldsymbol{\mu} = c(1, ..., 1)^T \in \boldsymbol{R_N}$.

Appendix

Throughout this section, we assume that $b_1 = \cdots = b_k = 1$, without loss of generality. Throughout, let $n_* = \sum_{i=1}^{k} n_i$. Let $\boldsymbol{y}_j = (\boldsymbol{x}_{1j} - \boldsymbol{\mu}_1)/n_1$ for $j = 1, ..., n_1$, and $\boldsymbol{y}_{j+\sum_{i'=1}^{i-1} n_{i'}} = (\boldsymbol{x}_{ij} - \boldsymbol{\mu}_i)/n_i$ for $j = 1, ..., n_i$ $(i \ge 2)$. Let $V_{n_* j} = \sum_{i=1}^{j-1} \boldsymbol{y}_i^T \boldsymbol{y}_j$ for $j = 2, ..., n_*$. Note that $E_{\boldsymbol{\theta}}\{(2 \sum_{j=2}^{n_*} V_{n_* j})^2\} = 2 \sum_{i=1}^{k} \text{tr}(\Sigma_i^2)(n_i - 1)/n_i^3 + 4 \sum_{i<j} \text{tr}(\Sigma_i \Sigma_j)/(n_i n_j)$ $(= K_v$, say).

Lemma 1 *Assume* (A-v) *and* (A-vi). *Then, it holds as* $p \to \infty$ *and* $n_i \to \infty$, $i = 1, ..., k$, *that*

$$\frac{4 \sum_{j=2}^{n_*} V_{n_* j}^2}{K_v} = 1 + o_p(1).$$

Proof Let $\boldsymbol{\Gamma}_i = (\boldsymbol{\gamma}_{i1}, ..., \boldsymbol{\gamma}_{ir_i})$, $i = 1, ..., k$. Let i_* be an integer such that $i \in [1 + \sum_{j=1}^{i_*-1} n_j, \sum_{j=1}^{i_*} n_j]$, where $\sum_{j=1}^{0} n_j = 0$. Note that $\mathrm{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j \boldsymbol{\Sigma}_{i'} \boldsymbol{\Sigma}_j) = \mathrm{tr}(\boldsymbol{\Sigma}_j^{1/2} \boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j^{1/2} \boldsymbol{\Sigma}_j^{1/2} \boldsymbol{\Sigma}_{i'} \boldsymbol{\Sigma}_j^{1/2}) \leq \{\mathrm{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j \boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j)\mathrm{tr}(\boldsymbol{\Sigma}_{i'} \boldsymbol{\Sigma}_j \boldsymbol{\Sigma}_{i'} \boldsymbol{\Sigma}_j)\}^{1/2}$. Under (A-v), we have for $i \neq i' \neq j$ that

$$n_{i_*}^2 n_{i'_*}^2 n_{j_*}^4 E_{\boldsymbol{\theta}}\{(\boldsymbol{y}_i^T \boldsymbol{y}_j)^2 (\boldsymbol{y}_{i'}^T \boldsymbol{y}_j)^2\} = n_{j_*}^4 E_{\boldsymbol{\theta}}(\boldsymbol{y}_j^T \boldsymbol{\Sigma}_{i_*} \boldsymbol{y}_j \boldsymbol{y}_j^T \boldsymbol{\Sigma}_{i'_*} \boldsymbol{y}_j)$$

$$= E_{\boldsymbol{\theta}}\left(\sum_{s,t,u,v}^{r_{j_*}} \boldsymbol{\gamma}_{j_*s}^T \boldsymbol{\Sigma}_{i_*} \boldsymbol{\gamma}_{j_*t} \boldsymbol{\gamma}_{j_*u}^T \boldsymbol{\Sigma}_{i'_*} \boldsymbol{\gamma}_{j_*v} w_{j_*sj} w_{j_*tj} w_{j_*uj} w_{j_*vj} \right)$$

$$= \mathrm{tr}(\boldsymbol{\Sigma}_{i_*} \boldsymbol{\Sigma}_{j_*})\mathrm{tr}(\boldsymbol{\Sigma}_{i'_*} \boldsymbol{\Sigma}_{j_*}) + 2\mathrm{tr}(\boldsymbol{\Sigma}_{i_*} \boldsymbol{\Sigma}_{j_*} \boldsymbol{\Sigma}_{i'_*} \boldsymbol{\Sigma}_{j_*})$$

$$+ O\left\{ \sum_{s=1}^{r_{j_*}} \boldsymbol{\gamma}_{j_*s}^T \boldsymbol{\Sigma}_{i_*} \boldsymbol{\gamma}_{j_*s} \boldsymbol{\gamma}_{j_*s}^T \boldsymbol{\Sigma}_{i'_*} \boldsymbol{\gamma}_{j_*s} \right\}$$

$$= \mathrm{tr}(\boldsymbol{\Sigma}_{i_*} \boldsymbol{\Sigma}_{j_*})\mathrm{tr}(\boldsymbol{\Sigma}_{i'_*} \boldsymbol{\Sigma}_{j_*}) + O\left[\{\mathrm{tr}(\boldsymbol{\Sigma}_{i_*} \boldsymbol{\Sigma}_{j_*} \boldsymbol{\Sigma}_{i_*} \boldsymbol{\Sigma}_{j_*})\mathrm{tr}(\boldsymbol{\Sigma}_{i'_*} \boldsymbol{\Sigma}_{j_*} \boldsymbol{\Sigma}_{i'_*} \boldsymbol{\Sigma}_{j_*})\}^{1/2}\right] \quad (14)$$

from the fact that

$$\sum_{s=1}^{r_{j_*}} \boldsymbol{\gamma}_{j_*s}^T \boldsymbol{\Sigma}_{i_*} \boldsymbol{\gamma}_{j_*s} \boldsymbol{\gamma}_{j_*s}^T \boldsymbol{\Sigma}_{i'_*} \boldsymbol{\gamma}_{j_*s}$$

$$\leq \left\{ \sum_{s=1}^{r_{j_*}} (\boldsymbol{\gamma}_{j_*s}^T \boldsymbol{\Sigma}_{i_*} \boldsymbol{\gamma}_{j_*s})^2 \sum_{s'=1}^{r_{j_*}} (\boldsymbol{\gamma}_{j_*s'}^T \boldsymbol{\Sigma}_{i'_*} \boldsymbol{\gamma}_{j_*s'})^2 \right\}^{1/2}$$

$$\leq \left\{ \sum_{s,t}^{r_{j_*}} (\boldsymbol{\gamma}_{j_*s}^T \boldsymbol{\Sigma}_{i_*} \boldsymbol{\gamma}_{j_*t})^2 \sum_{s',t'}^{r_{j_*}} (\boldsymbol{\gamma}_{j_*s'}^T \boldsymbol{\Sigma}_{i'_*} \boldsymbol{\gamma}_{j_*t'})^2 \right\}^{1/2}$$

$$= \{\mathrm{tr}(\boldsymbol{\Sigma}_{i_*} \boldsymbol{\Sigma}_{j_*} \boldsymbol{\Sigma}_{i_*} \boldsymbol{\Sigma}_{j_*})\mathrm{tr}(\boldsymbol{\Sigma}_{i'_*} \boldsymbol{\Sigma}_{j_*} \boldsymbol{\Sigma}_{i'_*} \boldsymbol{\Sigma}_{j_*})\}^{1/2}. \quad (15)$$

On the other hand, from Eq. 15, under (A-v), we have for $i \neq j$ that

$$E_{\boldsymbol{\theta}}\{(\boldsymbol{y}_i^T \boldsymbol{y}_j)^4\} n_{i_*}^4 n_{j_*}^4$$

$$= E_{\boldsymbol{\theta}}\left\{ \left(\sum_{s,u}^{r_{i_*}} \sum_{t,v}^{r_{j_*}} \boldsymbol{\gamma}_{i_*s}^T \boldsymbol{\gamma}_{j_*t} w_{i_*si} w_{j_*tj} \boldsymbol{\gamma}_{i_*u}^T \boldsymbol{\gamma}_{j_*v} w_{i_*ui} w_{j_*vj} \right)^2 \right\}$$

$$= \mathrm{tr}(\boldsymbol{\Sigma}_{i_*} \boldsymbol{\Sigma}_{j_*})^2 + E_{\boldsymbol{\theta}}\left\{ \left(\sum_{s \neq u}^{r_{i_*}} \sum_{t \neq v}^{r_{j_*}} \boldsymbol{\gamma}_{i_*s}^T \boldsymbol{\gamma}_{j_*t} w_{i_*si} w_{j_*tj} \boldsymbol{\gamma}_{i_*u}^T \boldsymbol{\gamma}_{j_*v} w_{i_*ui} w_{j_*vj} \right)^2 \right\}$$

$$+ O\{\mathrm{tr}(\boldsymbol{\Sigma}_{i_*} \boldsymbol{\Sigma}_{j_*} \boldsymbol{\Sigma}_{i_*} \boldsymbol{\Sigma}_{j_*})\} + O\left(\sum_{s=1}^{r_{j_*}} \boldsymbol{\gamma}_{j_*s}^T \boldsymbol{\Sigma}_{i_*} \boldsymbol{\gamma}_{j_*s} \boldsymbol{\gamma}_{j_*s}^T \boldsymbol{\Sigma}_{i_*} \boldsymbol{\gamma}_{j_*s} \right)$$

$$+ O\left(\sum_{s=1}^{r_{i_*}} \boldsymbol{\gamma}_{i_*s}^T \boldsymbol{\Sigma}_{j_*} \boldsymbol{\gamma}_{i_*s} \boldsymbol{\gamma}_{i_*s}^T \boldsymbol{\Sigma}_{j_*} \boldsymbol{\gamma}_{i_*s} \right)$$

$$= 3\mathrm{tr}(\boldsymbol{\Sigma}_{i_*} \boldsymbol{\Sigma}_{j_*})^2 + O\{\mathrm{tr}(\boldsymbol{\Sigma}_{i_*} \boldsymbol{\Sigma}_{j_*} \boldsymbol{\Sigma}_{i_*} \boldsymbol{\Sigma}_{j_*})\} = O\{\mathrm{tr}(\boldsymbol{\Sigma}_{i_*} \boldsymbol{\Sigma}_{j_*})^2\} \quad (16)$$

from the fact that $\text{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j) = \text{tr}\{(\boldsymbol{\Sigma}_i^{1/2}\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}_i^{1/2})^2\} \le \text{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j)^2$. Let $\boldsymbol{B}_j = \boldsymbol{\Sigma}_{j'}/n_{j'}^2$ for $j \in [1 + \sum_{i=1}^{j'-1} n_i, \sum_{i=1}^{j'} n_i]$, where $\sum_{i=1}^{0} n_i = 0$. Note that $E_{\boldsymbol{\theta}}(V_{n_*j}^2) = \sum_{i=1}^{j-1} \text{tr}(\boldsymbol{B}_i\boldsymbol{B}_j)$ for $j = 2, ..., n_*$, and $4\sum_{j=2}^{n_*}\sum_{i=1}^{j-1} \text{tr}(\boldsymbol{B}_i\boldsymbol{B}_j) = K_v$. From Eqs. 14–16, we have for $2 \le j < j'$ that

$$E_{\boldsymbol{\theta}}\big[\{V_{n_*j}^2 - E_{\boldsymbol{\theta}}(V_{n_*j}^2)\}^2\big] \le E_{\boldsymbol{\theta}}(V_{n_*j}^4) = O\left\{\sum_{i,i'}^{j-1} \text{tr}(\boldsymbol{B}_i\boldsymbol{B}_j)\text{tr}(\boldsymbol{B}_{i'}\boldsymbol{B}_j)\right\}; \qquad (17)$$

$$E_{\boldsymbol{\theta}}\big[\{V_{n_*j}^2 - E_{\boldsymbol{\theta}}(V_{n_*j}^2)\}\{V_{n_*j'}^2 - E_{\boldsymbol{\theta}}(V_{n_*j'}^2)\}\big]$$

$$= O\left[\sum_{i,i'}^{j-1}\{\text{tr}(\boldsymbol{B}_i\boldsymbol{B}_j\boldsymbol{B}_i\boldsymbol{B}_j)\text{tr}(\boldsymbol{B}_{i'}\boldsymbol{B}_{j'}\boldsymbol{B}_{i'}\boldsymbol{B}_{j'})\}^{1/2}\right]$$

$$+ O\left[\sum_{i=1}^{j-1}\text{tr}(\boldsymbol{B}_i\boldsymbol{B}_j)\{\text{tr}(\boldsymbol{B}_i\boldsymbol{B}_{j'}) + \text{tr}(\boldsymbol{B}_j\boldsymbol{B}_{j'})\}\right] \qquad (18)$$

from the fact that

$$E_{\boldsymbol{\theta}}\{(\boldsymbol{y}_i^T\boldsymbol{y}_j)^2(\boldsymbol{y}_i^T\boldsymbol{y}_{j'})(\boldsymbol{y}_j^T\boldsymbol{y}_{j'})\} \le \big[E_{\boldsymbol{\theta}}\{(\boldsymbol{y}_i^T\boldsymbol{y}_j)^4\}E_{\boldsymbol{\theta}}\{(\boldsymbol{y}_i^T\boldsymbol{y}_{j'})^2(\boldsymbol{y}_j^T\boldsymbol{y}_{j'})^2\}\big]^{1/2}$$

$$= O\big[\text{tr}(\boldsymbol{B}_i\boldsymbol{B}_j)\{\text{tr}(\boldsymbol{B}_i\boldsymbol{B}_{j'})\text{tr}(\boldsymbol{B}_j\boldsymbol{B}_{j'})\}^{1/2}\big]$$

$$= O\big[\text{tr}(\boldsymbol{B}_i\boldsymbol{B}_j)\{\text{tr}(\boldsymbol{B}_i\boldsymbol{B}_{j'}) + \text{tr}(\boldsymbol{B}_j\boldsymbol{B}_{j'})\}\big]$$

for $i < j < j'$. Then, from Eq. 17, we can obtain as $n_i \to \infty$, $i = 1, ..., k$, that

$$\sum_{j=2}^{n_*} E_{\boldsymbol{\theta}}\big[\{V_{n_*j}^2 - E_{\boldsymbol{\theta}}(V_{n_*j}^2)\}^2\big] \le \sum_{j=2}^{n_*} E_{\boldsymbol{\theta}}(V_{n_*j}^4) = O\left(\frac{K_v^2}{\min_{i=1,...,k} n_i}\right) = o(K_v^2). \qquad (19)$$

On the other hand, under (A-vi), we have for all i, i', j, j' that

$$\frac{\{\text{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j)\text{tr}(\boldsymbol{\Sigma}_{i'}\boldsymbol{\Sigma}_{j'}\boldsymbol{\Sigma}_{i'}\boldsymbol{\Sigma}_{j'})\}^{1/2}}{\{\text{tr}(\boldsymbol{\Sigma}_i^2)\text{tr}(\boldsymbol{\Sigma}_j^2)\text{tr}(\boldsymbol{\Sigma}_{i'}^2)\text{tr}(\boldsymbol{\Sigma}_{j'}^2)\}^{1/2}} \le \frac{\{\text{tr}(\boldsymbol{\Sigma}_i^4)\text{tr}(\boldsymbol{\Sigma}_j^4)\text{tr}(\boldsymbol{\Sigma}_{i'}^4)\text{tr}(\boldsymbol{\Sigma}_{j'}^4)\}^{1/4}}{\{\text{tr}(\boldsymbol{\Sigma}_i^2)\text{tr}(\boldsymbol{\Sigma}_j^2)\text{tr}(\boldsymbol{\Sigma}_{i'}^2)\text{tr}(\boldsymbol{\Sigma}_{j'}^2)\}^{1/2}}$$

$$\le \frac{\{\lambda_{i1}\lambda_{j1}\lambda_{i'1}\lambda_{j'1}\}^{1/2}}{\{\text{tr}(\boldsymbol{\Sigma}_i^2)\text{tr}(\boldsymbol{\Sigma}_j^2)\text{tr}(\boldsymbol{\Sigma}_{i'}^2)\text{tr}(\boldsymbol{\Sigma}_{j'}^2)\}^{1/4}}$$

$$\to 0$$

from the facts that $\text{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j) \le \text{tr}(\boldsymbol{\Sigma}_i^2\boldsymbol{\Sigma}_j^2) \le \text{tr}(\boldsymbol{\Sigma}_i^4)^{1/2}\text{tr}(\boldsymbol{\Sigma}_j^4)^{1/2}$ and $\text{tr}(\boldsymbol{\Sigma}_i^4)^{1/2} \le \lambda_{i1}\text{tr}(\boldsymbol{\Sigma}_i^2)^{1/2}$. Then, it holds under (A-vi) that

$$\frac{\{\text{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j)\text{tr}(\boldsymbol{\Sigma}_{i'}\boldsymbol{\Sigma}_{j'}\boldsymbol{\Sigma}_{i'}\boldsymbol{\Sigma}_{j'})\}^{1/2}/(n_in_jn_{i'}n_{j'})}{K_v^2}$$

$$= O\left[\frac{\{\text{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j)\text{tr}(\boldsymbol{\Sigma}_{i'}\boldsymbol{\Sigma}_{j'}\boldsymbol{\Sigma}_{i'}\boldsymbol{\Sigma}_{j'})\}^{1/2}/(n_in_jn_{i'}n_{j'})}{\{\text{tr}(\boldsymbol{\Sigma}_i^2)/n_i^2 + \text{tr}(\boldsymbol{\Sigma}_j^2)/n_j^2\}\{\text{tr}(\boldsymbol{\Sigma}_{i'}^2)/n_{i'}^2 + \text{tr}(\boldsymbol{\Sigma}_{j'}^2)/n_{j'}^2\}}\right]$$

$$= O\left[\frac{\{\text{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j)\text{tr}(\boldsymbol{\Sigma}_{i'}\boldsymbol{\Sigma}_{j'}\boldsymbol{\Sigma}_{i'}\boldsymbol{\Sigma}_{j'})\}^{1/2}}{\{\text{tr}(\boldsymbol{\Sigma}_i^2)\text{tr}(\boldsymbol{\Sigma}_j^2)\text{tr}(\boldsymbol{\Sigma}_{i'}^2)\text{tr}(\boldsymbol{\Sigma}_{j'}^2)\}^{1/2}}\right] \to 0.$$

Hence, from Eq. 18, under (A-vi), we can claim that

$$\sum_{2 \leq j < j'}^{n_*} E_\theta\big[\{V_{n_*j}^2 - E_\theta(V_{n_*j}^2)\}\{V_{n_*j'}^2 - E_\theta(V_{n_*j'}^2)\}\big] = o(K_v^2). \tag{20}$$

Thus by combining Eqs. 19 and 20, we have as $p \to \infty$ and $n_i \to \infty$, $i = 1, ..., k$, that

$$\mathrm{Var}_\theta\left(\sum_{j=2}^{n_*} V_{n_*j}^2\right) = o(K_v^2)$$

under (A-v) and (A-vi). Thus by using Chebyshev's inequality, from the fact that $4\sum_{j=2}^{n_*} E_\theta(V_{n_*j}^2) = K_v$, it holds that $4\sum_{j=2}^{n_*} V_{n_*j}^2 / K_v = 1 + o_p(1)$. Thus it concludes the result.   □

Lemma 2 *Let $I(\cdot)$ be the indicator function. For Lindeberg's condition, under* (A-v), *it holds as $p \to \infty$ and $n_i \to \infty$, $i = 1, ..., k$, that*

$$\sum_{j=2}^{n_*} E_\theta\left\{\frac{V_{n_*j}^2}{K_v} I\left(\frac{V_{n_*j}^2}{K_v} > \tau\right)\right\} \to 0$$

for any $\tau > 0$.

Proof By using Chebyshev's inequality and Schwarz's inequality, we have that

$$\sum_{j=2}^{n_*} E_\theta\left\{\frac{V_{n_*j}^2}{K_v} I\left(\frac{V_{n_*j}^2}{K_v} > \tau\right)\right\} \leq \sum_{j=2}^{n_*}\left[E_\theta\left(\frac{V_{n_*j}^4}{K_v^2}\right) E_\theta\left\{I\left(\frac{V_{n_*j}^2}{K_v} > \tau\right)\right\}\right]^{1/2}$$

$$\leq \tau^{-1} K_v^{-2} \sum_{j=2}^{n_*} E_\theta(V_{n_*j}^4). \tag{21}$$

Then, by combining Eq. 21 with Eq. 19, we can conclude the result.   □

Lemma 3 *Assume* (A-v). *Then, it holds as $p \to \infty$ and $n_i \to \infty$, $i = 1, ..., k$, that*

$$\frac{\widehat{K}}{K} = 1 + o_p(1).$$

Proof From Eqs. 14 and 16, we have for $i \neq j$ that

$$E_\theta\big[\{((x_{is} - \mu_i)^T(x_{jt} - \mu_j))^2 - \mathrm{tr}(\Sigma_i\Sigma_j)\}^2\big] = O\{\mathrm{tr}(\Sigma_i\Sigma_j)^2\};$$

$$E_\theta\big[\{(x_{is} - \mu_i)^T(x_{jt} - \mu_j)\}^2\{(x_{is'} - \mu_i)^T(x_{jt} - \mu_j)^2 - \mathrm{tr}(\Sigma_i\Sigma_j)^2\}\big]$$

$$= O\{\mathrm{tr}(\Sigma_i\Sigma_j\Sigma_i\Sigma_j)\} \quad (s \neq s') \tag{22}$$

under (A-v). We write that

$$\mathrm{tr}(S_{in_i}S_{jn_j}) = \sum_{s=1}^{n_i}\sum_{t=1}^{n_j}\frac{\big[\{x_{is} - \mu_i - (\overline{x}_{in_i} - \mu_i)\}^T\{x_{jt} - \mu_j - (\overline{x}_{jn_j} - \mu_j)\}\big]^2}{(n_i - 1)(n_j - 1)}.$$

Then, from Eq. 22, it holds as $p \to \infty$ and $n_i \to \infty$, $i = 1, ..., k$, that

$$\text{Var}_{\boldsymbol{\theta}}\{\text{tr}(\boldsymbol{S}_{in_i}\boldsymbol{S}_{jn_j})\} = o\{\text{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j)^2\} \tag{23}$$

for $i < j$. By combining Eq. 5 with Eq. 23, we can claim under (A-v) that $\widehat{K} = K\{1 + o_p(1)\}$ as $p \to \infty$ and $n_i \to \infty$, $i = 1, ..., k$. Thus it concludes the result.
□

Proofs of Theorem 4 and Corollary 4 Define $S_{n_*t} = \sum_{j=2}^{t} V_{n_*j}$ for $t = 2, ..., n_*$. Let $\mathcal{F}_{n_*i} = \sigma(\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_i)$ be the σ algebra by $\{\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_i\}$ for $i \geq 2$. Note that S_{n_*t} is of zero mean and square integrable. Note that $E_{\boldsymbol{\theta}}(S_{n_*t}|\mathcal{F}_i) = \sum_{j=2}^{i} V_{n_*j} = S_{n_*i}$ for $t > i$. Thus $\{S_{n_*i}, \mathcal{F}_{n_*i}\}_{i=2}^{n_*}$ is a sequence of zero mean and a square integrable martingale. We consider applying the martingale central limit theorem given by McLeish (1974). Refer to Section 2.6 in Ghosh et al. (1997) for the details of the martingale central limit theorem. We have as $n_i \to \infty$, $i = 1, ..., k$, that

$$||\boldsymbol{T_n} - \boldsymbol{\mu}||^2 - \widehat{\Sigma}_{\mathbf{n}}$$

$$= 2\sum_{i=1}^{k}\sum_{s<t} \frac{(\boldsymbol{x}_{is} - \boldsymbol{\mu})^T(\boldsymbol{x}_{it} - \boldsymbol{\mu})}{n_i(n_i - 1)} + 2\sum_{i<j}\sum_{s=1}^{n_i}\sum_{t=1}^{n_j} \frac{(\boldsymbol{x}_{is} - \boldsymbol{\mu})^T(\boldsymbol{x}_{jt} - \boldsymbol{\mu})}{n_in_j}$$

$$= 2\sum_{j=2}^{n_*} V_{n_*j} + 2\sum_{i=1}^{k}\sum_{s<t} \frac{(\boldsymbol{x}_{is} - \boldsymbol{\mu})^T(\boldsymbol{x}_{it} - \boldsymbol{\mu})}{n_i^2(n_i - 1)} = 2\sum_{j=2}^{n_*} V_{n_*j} + o_p(K^{1/2}). \tag{24}$$

Note that $K/K_v \to 1$ as $n_i \to \infty$, $i = 1, ..., k$. Then, by using the martingale central limit theorem, we obtain that

$$\frac{||\boldsymbol{T_n} - \boldsymbol{\mu}||^2 - \widehat{\Sigma}_{\mathbf{n}}}{K} = \frac{2\sum_{j=2}^{n_*} V_{n_*j}}{K_v} + o_p(1) \Rightarrow N(0, 1) \tag{25}$$

by combining Lemmas 1 and 2 with Eq. 24. Thus it concludes the result of Theorem 4. On the other hand, by using Lemma 3, we obtain from Eq. 25 that $(||\boldsymbol{T_n} - \boldsymbol{\mu}||^2 - \widehat{\Sigma}_{\mathbf{n}})/\widehat{K} \Rightarrow N(0, 1)$. It concludes the result of Corollary 4.
□

Proofs of Theorem 5 and Corollary 5 Note that

$$\widehat{T}_{\mathbf{n}} - ||\boldsymbol{\mu}||^2 = ||\boldsymbol{T_n} - \boldsymbol{\mu}||^2 - \widehat{\Sigma}_{\mathbf{n}} + 2(\boldsymbol{T_n} - \boldsymbol{\mu})^T\boldsymbol{\mu}$$

$$= ||\boldsymbol{T_n} - \boldsymbol{\mu}||^2 - \widehat{\Sigma}_{\mathbf{n}} + o_p(K^{1/2})$$

and $K_*/K \to 1$ under (A-vii). Thus in a way similar to the proofs of Theorem 4 and Corollary 4, we can conclude the results.
□

Proof of Corollary 2 Note that $\sum_{i=1}^{k} \lambda_{i1}/(n_iK^{1/2}) = O\{\sum_{i=1}^{k} \lambda_{i1}/\text{tr}(\boldsymbol{\Sigma}_i^2)^{1/2}\} = o(1)$ under (A-vi). Then, it holds under (A-vi) that

$$\sum_{i=1}^{k} \frac{\boldsymbol{\mu}^T\boldsymbol{\Sigma}_i\boldsymbol{\mu}}{n_iK} \leq ||\boldsymbol{\mu}||^2 \sum_{i=1}^{k} \frac{\lambda_{i1}}{n_iK} = \frac{||\boldsymbol{\mu}||^2}{K^{1/2}} \sum_{i=1}^{k} \frac{\lambda_{i1}}{n_iK^{1/2}} = \frac{||\boldsymbol{\mu}||^2}{K^{1/2}} \times o(1),$$

so that we obtain that $K/||\boldsymbol{\mu}||^4 = o(1)$ under (A-vi) and (A-viii). Then, it holds that $\sum_{i=1}^{k} \boldsymbol{\mu}^T \boldsymbol{\Sigma}_i \boldsymbol{\mu}/(n_i||\boldsymbol{\mu}||^4) \leq \sum_{i=1}^{k} \lambda_{i1}/(n_i||\boldsymbol{\mu}||^2) \leq \sum_{i=1}^{k} \mathrm{tr}(\boldsymbol{\Sigma}_i^2)^{1/2} /(n_i||\boldsymbol{\mu}||^2) = O(K^{1/2}/||\boldsymbol{\mu}||^2) = o(1)$. Thus we have that $\mathrm{Var}_{\boldsymbol{\theta}}(\widehat{T}_{\mathbf{n}})/||\boldsymbol{\mu}||^4 = K_*/||\boldsymbol{\mu}||^4 = o(1)$ under (A-vi) and (A-viii). It concludes the result.   □

Proof of Corollary 3 By using Schwarz's inequality, it holds under (A-v') that $E_{\boldsymbol{\theta}}[\{(\boldsymbol{x}_{il} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_j(\boldsymbol{x}_{il} - \boldsymbol{\mu}_i) - \mathrm{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j)\}\{(\boldsymbol{x}_{il} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_{j'}(\boldsymbol{x}_{il} - \boldsymbol{\mu}_i) - \mathrm{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_{j'})\}] \leq [\mathrm{Var}_{\boldsymbol{\theta}}\{(\boldsymbol{x}_{il} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_j(\boldsymbol{x}_{il} - \boldsymbol{\mu}_i)\}\mathrm{Var}_{\boldsymbol{\theta}}\{(\boldsymbol{x}_{il} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_{j'}(\boldsymbol{x}_{il} - \boldsymbol{\mu}_i)\}]^{1/2} = O[\{\mathrm{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j \boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j) \times \mathrm{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_{j'} \boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_{j'})\}^{1/2}]$ for $i, j, j' = 1, ..., k$. Then, we have under (A-v') that

$$n_{j_*}^4 E_{\boldsymbol{\theta}}\{\boldsymbol{y}_j^T \boldsymbol{\Sigma}_{i_*} \boldsymbol{y}_j \boldsymbol{y}_j^T \boldsymbol{\Sigma}_{i_*'} \boldsymbol{y}_j\} - \mathrm{tr}(\boldsymbol{\Sigma}_{i_*} \boldsymbol{\Sigma}_{j_*})\mathrm{tr}(\boldsymbol{\Sigma}_{i_*'} \boldsymbol{\Sigma}_{j_*})$$

$$= E_{\boldsymbol{\theta}}[\{n_{j_*}^2 \boldsymbol{y}_j^T \boldsymbol{\Sigma}_{i_*} \boldsymbol{y}_j - \mathrm{tr}(\boldsymbol{\Sigma}_{i_*} \boldsymbol{\Sigma}_{j_*})\}\{n_{j_*}^2 \boldsymbol{y}_j^T \boldsymbol{\Sigma}_{i_*'} \boldsymbol{y}_j - \mathrm{tr}(\boldsymbol{\Sigma}_{i_*'} \boldsymbol{\Sigma}_{j_*})\}]$$

$$= O[\{\mathrm{tr}(\boldsymbol{\Sigma}_{i_*} \boldsymbol{\Sigma}_{j_*} \boldsymbol{\Sigma}_{i_*} \boldsymbol{\Sigma}_{j_*})\mathrm{tr}(\boldsymbol{\Sigma}_{i_*'} \boldsymbol{\Sigma}_{j_*} \boldsymbol{\Sigma}_{i_*'} \boldsymbol{\Sigma}_{j_*})\}^{1/2}],$$

so that Eq. 14 holds. On the other hand, under (A-v'), it holds for $i \neq j$ that $n_{i_*}^4 n_{j_*}^4 E_{\boldsymbol{\theta}}\{(\boldsymbol{y}_i^T \boldsymbol{y}_j)^4\} = \mathrm{Var}_{\boldsymbol{\theta}}\{n_{i_*}^2 n_{j_*}^2 (\boldsymbol{y}_i^T \boldsymbol{y}_j)^2\} + \mathrm{tr}(\boldsymbol{\Sigma}_{i_*} \boldsymbol{\Sigma}_{j_*})^2 = O\{\mathrm{tr}(\boldsymbol{\Sigma}_{i_*} \boldsymbol{\Sigma}_{j_*})^2\}$, so that Eq. 16 holds. Thus we claim Lemmas 1 and 2 after replacing (A-v) with (A-v'). We can conclude the result in a way similar to the proofs of Theorems 4 and 5.   □

Proof of Theorem 7 We have from Eq. 10 that $K^{1/2} \leq \delta/z_{\alpha/2}$. Then, from Theorem 4, we claim as $p \to \infty$ that

$$P_{\boldsymbol{\theta}}(\boldsymbol{\mu} \in \boldsymbol{R}_{\mathbf{n}}) = P_{\boldsymbol{\theta}}(\max\{-\delta + \widehat{\Sigma}_{\mathbf{n}}, \, 0\} \leq ||\boldsymbol{T}_{\mathbf{n}} - \boldsymbol{\mu}||^2 \leq \delta + \widehat{\Sigma}_{\mathbf{n}})$$

$$= P_{\boldsymbol{\theta}}\left(\frac{-\delta}{K^{1/2}} \leq \frac{||\boldsymbol{T}_{\mathbf{n}} - \boldsymbol{\mu}||^2 - \widehat{\Sigma}_{\mathbf{n}}}{K^{1/2}} \leq \frac{\delta}{K^{1/2}}\right)$$

$$= P_{\boldsymbol{\theta}}\left(\left|\frac{||\boldsymbol{T}_{\mathbf{n}} - \boldsymbol{\mu}||^2 - \widehat{\Sigma}_{\mathbf{n}}}{K^{1/2}}\right| \leq \frac{\delta}{K^{1/2}}\right)$$

$$\geq P_{\boldsymbol{\theta}}\left(\left|\frac{||\boldsymbol{T}_{\mathbf{n}} - \boldsymbol{\mu}||^2 - \widehat{\Sigma}_{\mathbf{n}}}{K^{1/2}}\right| \leq z_{\alpha/2}\right)$$

$$= P_{\boldsymbol{\theta}}(|N(0, 1)| < z_{\alpha/2}) + o(1) = 1 - \alpha + o(1) \qquad (26)$$

under (A-v) and (A-vi). Thus the proof is completed.   □

Proof of Theorem 8 From Eqs. 5 and 12, it holds as $p \to \infty$ that $|N_i - C_i| = o_p(C_i^{1/2})$ under (A-v) and (A-vi). Then, we write that $|N_i - C_i| = O_p(\omega C_i^{1/2})$, where $\omega \, (> 0)$ is a variable such that $\omega \to 0$ as $p \to \infty$. Let $C_{iL} = \lfloor C_i - (\omega C_i)^{1/2} \rfloor$, $i = 1, ..., k$. We claim as $p \to \infty$ that $\max\{m_i, C_{iL}\} \leq N_i < C_i + (\omega C_i)^{1/2}$ w.p.1. Then, in a way similar to the proofs of Theorems 2.4 and 2.5 in Aoshima and Yata (2011a), we claim that

$$||\boldsymbol{T}_{\mathbf{N}} - \boldsymbol{\mu}||^2 - \widehat{\Sigma}_{\mathbf{N}} = ||\boldsymbol{T}_{\mathbf{C}_L} - \boldsymbol{\mu}||^2 - \widehat{\Sigma}_{\mathbf{C}_L} + o_p(\delta),$$

where $\boldsymbol{C}_L = (C_{1L}, ..., C_{kL})$. From the fact that $C_{iL}/C_i \to 1$ as $p \to \infty$, similarly to Eq. 26, we can obtain under (A-v) and (A-vi) that

$$P_{\boldsymbol{\theta}}(\boldsymbol{\mu} \in \boldsymbol{R}_{\mathbf{N}}) = P_{\boldsymbol{\theta}}\left(\left|||\boldsymbol{T}_{\mathbf{N}} - \boldsymbol{\mu}||^2 - \widehat{\Sigma}_{\mathbf{N}}\right| \leq \delta\right)$$

$$= P_{\boldsymbol{\theta}}\left(\left|||\boldsymbol{T}_{\mathbf{C}_L} - \boldsymbol{\mu}||^2 - \widehat{\Sigma}_{\mathbf{C}_L}\right| \leq \delta\right) + o(1) \geq 1 - \alpha + o(1).$$

It concludes the result.   □

References

Aoshima M, Yata K (2011a) Two-stage procedures for high-dimensional data. Seq Anal 30:356–399 (Editor's special invited paper)

Aoshima M, Yata K (2011b) Authors' response. Seq Anal 30:432–440

Aoshima M, Yata K (2011c) Effective methodologies for statistical inference on microarray studies. In: Spiess PE (ed) Prostate cancer - from bench to bedside. InTech, pp 13–32

Bai Z, Sarandasa H (1996) Effect of high dimension: by an example of a two sample problem. Stat Sin 6:311–329

Chen SX, Qin YL (2010) A two-sample test for high-dimensional data with applications to gene-set testing. Ann Stat 38:808–835

Chiaretti S, Li X, Gentleman R, Vitale A, Vignetti M, Mandelli F, Ritz J, Foa R (2004) Gene expression profile of adult T-cell acute lymphocytic leukemia identifies distinct subsets of patients with different response to therapy and survival. Blood 103:2771–2778

Ghosh M, Mukhopadhyay N, Sen PK (1997) Sequential estimation. Wiley, New York

McLeish DL (1974) Dependent central limit theorems and invariance principles. Ann Probab 2:620–628

Pollard KS, Dudoit S, van der Laan MJ (2005) Multiple testing procedures: R multitest package and applications to genomics. In: Gentleman R, Carey V, Huber W, Irizarry R, Dudoit S (eds) Bioinformatics and computational biology solutions using R and bioconductor. Springer, New York, pp 249–271

Srivastava MS (2005) Some tests concerning the covariance matrix in high dimensional data. J Jpn Stat Soc 35:251–272

Yata K, Aoshima M (2010) Effective PCA for high-dimension, low-sample-size data with singular value decomposition of cross data matrix. J Multivar Anal 101:2060–2077

Yata K, Aoshima M (2012) Inference on high-dimensional mean vectors with fewer observations than the dimension. Methodol Comput Appl Probab 14:459–476

Yata K, Aoshima M (2013) Correlation tests for high-dimensional data using extended cross-data-matrix methodology. J Multivar Anal 117:313–331