Inference on High-Dimensional Mean Vectors with Fewer Observations Than the Dimension

Kazuyoshi Yata · Makoto Aoshima

Abstract We focus on inference about high-dimensional mean vectors when the sample size is much fewer than the dimension. Such data situation occurs in many areas of modern science such as genetic microarrays, medical imaging, text recognition, finance, chemometrics, and so on. First, we give a given-radius confidence region for mean vectors. This inference can be utilized as a variable selection of high-dimensional data. Next, we give a given-width confidence interval for squared norm of mean vectors. This inference can be utilized in a classification procedure of high-dimensional data. In order to assure a prespecified coverage probability, we propose a two-stage estimation methodology and determine the required sample size for each inference. Finally, we demonstrate how the new methodologies perform by using a microarray data set.

1 Introduction

Suppose we have independent and p-variate populations, π_i, $i = 1, ..., k$, having unknown mean vector $\boldsymbol{\mu}_i = (\mu_{i1}, ..., \mu_{ip})^T$ and unknown covariance matrix $\boldsymbol{\Sigma}_i(> \boldsymbol{O})$ for each i. *We do not assume that $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_k$.* The eigen-decomposition of $\boldsymbol{\Sigma}_i$ $(i = 1, ..., k)$ is $\boldsymbol{\Sigma}_i = \boldsymbol{H}_i \boldsymbol{\Lambda}_i \boldsymbol{H}_i^T$, where $\boldsymbol{\Lambda}_i$ is a diagonal matrix of eigenvalues $\lambda_{i1} \geq \cdots \geq \lambda_{ip} > 0$ and $\boldsymbol{H}_i = [\boldsymbol{h}_{i1}, ..., \boldsymbol{h}_{ip}]$ is

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

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

an orthogonal matrix of corresponding eigenvectors. Having recorded i.i.d. samples, $\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{in_i}$, from each π_i, we have a $p \times n_i$ $(p > n_i)$ data matrix $\boldsymbol{X}_i = [\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{in_i}]$, where $\boldsymbol{x}_{ij} = (x_{i1j}, ..., x_{ipj})^T$, $j = 1, ..., n_i$. Then, $\boldsymbol{Z}_i = \boldsymbol{\Lambda}_i^{-1/2} \boldsymbol{H}_i^T (\boldsymbol{X}_i - [\boldsymbol{\mu}_i, ..., \boldsymbol{\mu}_i])$ is a $p \times n_i$ sphered data matrix from a distribution with the identity covariance matrix. Here, we write $\boldsymbol{Z}_i = [\boldsymbol{z}_{i1}, ..., \boldsymbol{z}_{in_i}]$ and $\boldsymbol{z}_{ij} = (z_{i1j}, ..., z_{ipj})^T$, $j = 1, ..., n_i$. Note that $E(z_{ijl}^2) = 1$ and $E(z_{ijl}z_{ij'l}) = 0$ for $i = 1, ..., k$; $j (\neq j') = 1, ..., p$; $l = 1, ..., n_i$. We assume that $\lambda_{ip} > 0$ $(i = 1, ..., k)$ as $p \to \infty$ and the fourth moments of each variable in $\boldsymbol{Z}_i$ are uniformly bounded.

In this paper, we assume one of the following three assumptions for π_i's:

(A-i) $\pi_i : N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ for $i = 1, ..., k$;
(A-ii) $z_{ijl}, j = 1, ..., p$ are independent for $i = 1, ..., k$;
(A-iii) $E(z_{ijl}^2 z_{isl}^2) = 1$ and $E(z_{ijl}z_{isl}z_{itl}z_{iul}) = 0$, $j \neq s, t, u$, and $\{x_{ijl} - \mu_{ij}\}_{j \in N}$ is a strictly stationary sequence and ρ-mixing for $i = 1, ..., k$.

Note that (A-i) implies (A-ii). The concept of ρ-mixing was first developed by Kolmogorov and Rozanov (1960). See Bradley (2005) for a clear and insightful discussion. Throughout this paper, we assume the following conditions for $\boldsymbol{\Sigma}_i$'s:

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

We assume the following extra condition when applying (A-iii):

(A-v) $\dfrac{\text{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j)}{p} \to c_{ij}$ as $p \to \infty$ for all $i, j = 1, ..., k$, where c_{ij}'s are positive constants.

Remark 1. If all λ_{ij}'s are bounded, (A-iv) trivially holds. For a spiked model such as $\lambda_{ij} = a_{ij}p^{\alpha_{ij}}$ $(j = 1, ..., m_i)$ and $\lambda_{ij} = c_{ij}$ $(j = m_i + 1, ..., p)$ with positive constants a_{ij}'s, c_{ij}'s and α_{ij}'s, (A-iv) holds under the condition that $\alpha_{ij} < 1/2$ for $j = 1, ..., m_i(< \infty)$; $i = 1, ..., k$. See Yata and Aoshima (2009b, 2010a) for the details of a spiked model. In an actual data analysis, one may examine (A-iv) by using the cross-data-matrix methodology given by Yata and Aoshima (2010a) or the noise reduction methodology given by Yata and Aoshima (2011). As an interesting example, both (A-iv) and (A-v) hold for $\boldsymbol{\Sigma}_{i'} = c_{i'}(\rho_{i'}^{|i-j|^{q_{i'}}})$, $i' = 1, ..., k$, where $c_{i'}$'s, $q_{i'}$'s and $\rho_{i'}$'s(< 1) are positive constants.

Let $\boldsymbol{\mu} = \sum_{i=1}^k b_i \boldsymbol{\mu}_i$, where b_i's are known and nonzero scalars. Let us write that $\boldsymbol{T_n} = \sum_{i=1}^k b_i \overline{\boldsymbol{x}}_{in_i}$, where $\boldsymbol{n} = (n_1, ..., n_k)$ and $\overline{\boldsymbol{x}}_{in_i} = \sum_{j=1}^{n_i} \boldsymbol{x}_{ij}/n_i$. One choice of making inference on $\boldsymbol{\mu}$ is to construct a confidence region by $\boldsymbol{R_n} = \{\boldsymbol{\mu} \in R^p : ||\boldsymbol{T_n} - \boldsymbol{\mu}|| \leq d\}$, where $||\cdot||$ denotes the Euclidean norm. Let $\boldsymbol{\theta} = (\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_1, ..., \boldsymbol{\Sigma}_k)$ for given k. Then, the requirement is given by

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

for given and fixed d (> 0) and $\alpha \in (0,1)$. There are many literatures related to this problem when p is fixed less than n_i. One may refer to Ghosh et al. (1997), Aoshima and Mukhopadhyay (1998), Aoshima et al. (2002), Aoshima and Takada (2004), Aoshima (2005), Yata and Aoshima (2009a) and Aoshima and Yata (2010) among others in which Stein (1945)-type two-stage procedures were proposed in a typical multivariate context. Especially, Aoshima and Yata (2010) provided a general methodology to make a Stein-type two-stage procedure asymptotically second-order consistent for a variety of multivariate inference problems such as multiple comparisons and bioequivalence tests. For the concept of second-order efficiency, refer to Ghosh et al. (1997). In an extreme high-dimensional case, those methodologies tend to satisfy the probability requirement such as (1) excessively by taking overly samples. Yata (2010) created a new-type two-stage procedure that meets the equality in (1) approximately with a smaller sample size when p is extremely large. Note that $E(||\boldsymbol{T_n} - \boldsymbol{\mu}||^2) = \sum_{i=1}^{k} b_i^2 \mathrm{tr}(\boldsymbol{\Sigma}_i)$ ($= \Sigma_\mathbf{n}$, say). Thus $||\boldsymbol{T_n} - \boldsymbol{\mu}||^2$ behaves around a certain positive quantity, $\Sigma_\mathbf{n}$. Since it holds that $\Sigma_\mathbf{n} = O(\sum_{i=1}^{k} p/n_i)$ under (A-iv), one cannot claim (1) for a fixed span d ($< \infty$) when having $n_i = o(p)$, $i = 1, ..., k$, that are fewer observations than the dimension. To overcome this inconvenience, Aoshima and Yata (2011) have recently created a new confidence region called "given-bandwidth confidence region" in a high-dimension, low-sample-size context.

A common feature of high-dimensional data is that, while the data dimension is high, 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 n is the sample size. The HDLSS asymptotics, where only $p \to \infty$ while n is fixed, were studied by Hall et al. (2005), Ahn et al. (2007) and Yata and Aoshima (2012). They explored conditions to give a geometric representation of HDLSS data. The HDLSS asymptotics usually regulate either the population distribution by the normality or the dependency of the random variables in the sphered data matrix by a ρ-mixing condition. However, Yata and Aoshima (2010b) developed the HDLSS asymptotics without assuming either the normality or a ρ-mixing condition. Yata and Aoshima (2009b) succeeded in investigating the consistency properties of both eigenvalues and eigenvectors of the sample covariance matrix in more general settings that include the case when all eigenvalues are in the range of sphericity. In addition, Yata and Aoshima (2010a) developed the *cross-data-matrix methodology* that provides effective inference on PCA and clustering for HDLSS data. Recently, Aoshima and Yata (2011) have developed a variety of inference for HDLSS data such as a given-bandwidth confidence region, two-sample tests, classification, variable selection, regression, pathway analysis and so on.

In this paper, we focus on inference about high-dimensional mean vectors for HDLSS data. In Section 2, we give a given-radius confidence region for mean vectors. In Section 3, we give a given-width confidence interval for squared norm of mean vectors. In order to assure a prespecified coverage probability, we propose a two-stage estimation methodology and determine the re-

quired sample size for each inference. In Section 4, we demonstrate how the new methodologies perform by using a microarray data set.

2 Given-Radius Confidence Region for Mean Vectors

In this section, we consider a confidence region for the loss function, $||\boldsymbol{T_n} - \boldsymbol{\mu}||^2$, defined by

$$\boldsymbol{R_{n,W}} = \{\boldsymbol{\mu} \in R^p : ||\boldsymbol{T_n} - \boldsymbol{\mu}||^2 \leq W\}. \tag{2}$$

Aoshima and Yata (2011) showed that for a given confidence coefficient, $\boldsymbol{R_{n,W}}$ does not exist in the HDLSS context that $n_i/p \to 0$ as $p \to \infty$ if W is fixed free from p. We assume that $W = W(p) \to \infty$ as $p \to \infty$ and $W/\min_{1 \leq i \leq k} \mathrm{tr}(\boldsymbol{\Sigma}_i) = o(1)$. Our goal is to construct $\boldsymbol{R_{n,W}}$ satisfying

$$P_{\boldsymbol{\theta}}(\boldsymbol{\mu} \in \boldsymbol{R_{n,W}}) \geq 1 - \alpha \tag{3}$$

for given W (> 0) and $\alpha \in (0, 1/2)$.

2.1 Asymptotic Normality and Sample Size Determination

For the loss function $||\boldsymbol{T_n} - \boldsymbol{\mu}||^2$, Aoshima and Yata (2011) gave the following result.

Theorem 2.1(Aoshima and Yata, 2011). *Assume either (A-ii) or (A-iii) with (A-v). Then, we have as $p \to \infty$ and $n_i \to \infty$, $i = 1, ..., k$, that*

$$\frac{||\boldsymbol{T_n} - \boldsymbol{\mu}||^2 - \Sigma_{\mathbf{n}}}{\sqrt{2 \sum_{i,j} b_i^2 b_j^2 \, tr(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j)/(n_i n_j)}} \Rightarrow N(0, 1), \tag{4}$$

where "⇒" denotes the convergence in distribution and $N(0, 1)$ denotes a random variable distributed as the standard normal distribution.

From the fact that $\mathrm{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j) \leq \sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_i^2) \mathrm{tr}(\boldsymbol{\Sigma}_j^2)}$, it holds that

$$\sqrt{2 \sum_{i,j} b_i^2 b_j^2 \mathrm{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j)/(n_i n_j)} \leq \sum_{i=1}^{k} b_i^2 \sqrt{2 \, \mathrm{tr}(\boldsymbol{\Sigma}_i^2)}/n_i.$$

Let z_α be a constant such that $P(N(0, 1) > z_\alpha) = \alpha$. By using (4), we consider n_i's such that

$$\min \left((W/p) \sum_{i=1}^{k} n_i \right) \quad \text{subject to} \quad \frac{W - \Sigma_{\mathbf{n}}}{\sum_{i=1}^{k} b_i^2 \sqrt{2 \, \mathrm{tr}(\boldsymbol{\Sigma}_i^2)}/n_i} \geq z_\alpha$$

as $p \to \infty$. Then, we find the sample size for each π_i as

$$n_i \geq \frac{1}{W}|b_i|\sqrt{\text{tr}(\boldsymbol{\Sigma}_i)}\sum_{j=1}^{k}|b_j|\sqrt{\text{tr}(\boldsymbol{\Sigma}_j)} + \frac{z_\alpha\sqrt{2}}{W}|b_i|\sqrt{\text{tr}(\boldsymbol{\Sigma}_i)}\sum_{j=1}^{k}|b_j|\sqrt{\frac{\text{tr}(\boldsymbol{\Sigma}_j^2)}{\text{tr}(\boldsymbol{\Sigma}_j)}}$$
$$(= C_i, \text{ say}). \quad (5)$$

Note that $C_i = W^{-1}|b_i|\sqrt{\text{tr}(\boldsymbol{\Sigma}_i)}\sum_{j=1}^{k}|b_j|\sqrt{\text{tr}(\boldsymbol{\Sigma}_j)} + o(p/W) = o(p)$ under (A-iv). Thus it holds that $C_i/p \to 0$ as $p \to \infty$, $i = 1,...,k$. We have the following theorem.

Theorem 2.2. *Assume either (A-ii) or (A-iii) with (A-v). For n_i's satisfying (5), we have as $p \to \infty$ that*

$$\liminf P_{\boldsymbol{\theta}}(\boldsymbol{\mu} \in \boldsymbol{R}_{\mathbf{n},W}) \geq 1 - \alpha.$$

We emphasize that $\boldsymbol{R}_{\mathbf{n},W}$ meets requirement (3) in the HDLSS context that $n_i/p \to 0$ as $p \to \infty$.

Remark 2. One can claim Theorem 2.2 even for a constant W such that $W < \infty$ as $p \to \infty$. However, since one has that $0 < C_i/p < \infty$ in such a case, $\boldsymbol{R}_{\mathbf{n},W}$ cannot meet requirement (3) in the HDLSS context that $n_i/p \to 0$ as $p \to \infty$.

2.2 Two-stage procedure

Since $\boldsymbol{\Sigma}_i$'s are unknown, it is necessary to estimate C_i's in (5) with some pilot samples. Along the line of Mukhopadhyay and Duggan (1997, 1999), we suppose the following assumption: There exists a known and positive lower bound $\sigma_{i\star}$ for $\text{tr}(\boldsymbol{\Sigma}_i)$ such that $\sigma_{i\star}/\text{tr}(\boldsymbol{\Sigma}_i) \in (0,1)$, $i = 1,...,k$, as $p \to \infty$. We proceed the following two steps:

1. Let $\tau_\star = \min_{1 \leq i \leq k} |b_i|\sqrt{\sigma_{i\star}}\sum_{j=1}^{k}|b_j|\sqrt{\sigma_{j\star}}$. Having a fixed integer $m_0 (\geq 4)$, define

$$m = \max\left\{m_0, \left[\frac{\tau_\star}{W}\right] + 1\right\}, \quad (6)$$

where $[x]$ denotes the largest integer less than x. According to (6), take pilot samples $\boldsymbol{x}_{ij}$, $j = 1,...,m$, of size m from each π_i. Let $m_1 = [m/2] + 1$ and $m_2 = m - m_1$. Let $\overline{\boldsymbol{x}}_{im} = \sum_{j=1}^{m}\boldsymbol{x}_{ij}/m$, $\overline{\boldsymbol{x}}_{im(1)} = \sum_{j=1}^{m_1}\boldsymbol{x}_{ij}/m_1$ and $\overline{\boldsymbol{x}}_{im(2)} = \sum_{j=m_1+1}^{m}\boldsymbol{x}_{ij}/m_2$. Then, we calculate

$$\boldsymbol{S}_{im} = \frac{\sum_{j=1}^{m}(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_{im})(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_{im})^T}{m - 1},$$

$$\boldsymbol{S}_{im(1)} = \frac{\sum_{j=1}^{m_1}(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_{im(1)})(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_{im(1)})^T}{m_1 - 1}$$

$$\text{and} \quad \boldsymbol{S}_{im(2)} = \frac{\sum_{j=m_1+1}^{m}(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_{im(2)})(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_{im(2)})^T}{m_2 - 1} \quad (7)$$

for each π_i. Define the total sample size for each π_i by

$$
N_i = \max\Bigg\{ m, \ \Bigg[\frac{1}{W}|b_i|\sqrt{\mathrm{tr}(\boldsymbol{S}_{im})}\sum_{j=1}^{k}|b_j|\sqrt{\mathrm{tr}(\boldsymbol{S}_{jm})}
$$

$$
+ \frac{z_\alpha\sqrt{2}}{W}|b_i|\sqrt{\mathrm{tr}(\boldsymbol{S}_{im})}\sum_{j=1}^{k}|b_j|\sqrt{\frac{\mathrm{tr}(\boldsymbol{S}_{jm(1)}\boldsymbol{S}_{jm(2)})}{\mathrm{tr}(\boldsymbol{S}_{jm})}} \ \Bigg] + 1 \Bigg\}.
$$

(8)

Let $\boldsymbol{N} = (N_1, ..., N_k)$.

2. Take additional samples $\boldsymbol{x}_{ij}$, $j = m+1, ..., N_i$, of size $N_i - m$ from each π_i. By combining the initial samples and the additional samples, calculate $\overline{\boldsymbol{x}}_{iN_i} = \sum_{j=1}^{N_i}\boldsymbol{x}_{ij}/N_i$ for each π_i. Then, define $\boldsymbol{R}_{\mathbf{N},W}$ according to (2) with $\boldsymbol{T}_{\mathbf{N}} = \sum_{i=1}^{k}b_i\overline{\boldsymbol{x}}_{iN_i}$.

We have the following theorem.

Theorem 2.3. *Assume either (A-ii) or (A-iii) with (A-v). For the two-stage procedure given by (6)-(8), we have as $p \to \infty$ that*

$$
\liminf P_{\boldsymbol{\theta}}(\boldsymbol{\mu} \in \boldsymbol{R}_{\mathbf{N},W}) \geq 1 - \alpha.
$$

Remark 3. Assume either (A-ii) or (A-iii). It holds as $p \to \infty$ that $N_i/C_i = 1 + o_p(1)$, so that $N_i/p = o_p(1)$ for $i = 1, ..., k$. It should be noted that the result given by Theorem 2.3 can be claimed in a HDLSS situation.

Theorem 2.4. *Assume (A-i). For (6)-(8), it holds as $p \to \infty$ that*

$$
\limsup|E_{\boldsymbol{\theta}}(N_i - C_i)| \leq 1 \quad and \quad Var_{\boldsymbol{\theta}}(N_i) = O(1).
$$

Remark 4. We emphasize that the result given by Theorem 2.3 can be claimed as long as $\sigma_{i\star}/p > 0$ as $p \to \infty$ for $i = 1, ..., k$. In that sense, the two-stage procedure is quite robust for the misidentification of $\sigma_{i\star}$.

Remark 5. Yata (2010) considered an estimator of $\mathrm{tr}(\boldsymbol{\Sigma}_i^2)$ by $\mathrm{tr}(\boldsymbol{S}_{im(1)}\boldsymbol{S}_{im(2)})$. Note that $E_{\boldsymbol{\theta}}(\boldsymbol{S}_{im(1)}\boldsymbol{S}_{im(2)}) = \mathrm{tr}(\boldsymbol{\Sigma}_i^2)$ and $\mathrm{tr}(\boldsymbol{S}_{im(1)}\boldsymbol{S}_{im(2)}) \geq 0$ w.p.1. Under either (A-ii) or (A-iii), it holds as $p \to \infty$ and $m \to \infty$ that

$$
Var_{\boldsymbol{\theta}}\left(\frac{\mathrm{tr}(\boldsymbol{S}_{im(1)}\boldsymbol{S}_{im(2)})}{\mathrm{tr}(\boldsymbol{\Sigma}_i^2)}\right) = \frac{8}{m^2}(1 + o(1)) + O\left(\frac{\mathrm{tr}(\boldsymbol{\Sigma}_i^4)}{\mathrm{tr}(\boldsymbol{\Sigma}_i^2)^2 m}\right).
$$

On the other hand, Bai and Saranadasa (1996) and Srivastava (2005) considered an estimator of $\mathrm{tr}(\boldsymbol{\Sigma}_i^2)$ by $\mathrm{tr}(\widehat{\boldsymbol{\Sigma}_i^2}) = c_m^{-1}\{\mathrm{tr}(\boldsymbol{S}_{im}^2) - \mathrm{tr}(\boldsymbol{S}_{im})^2/(m-1)\}$ with

$c_m = (m-2)(m+1)/(m-1)^2$. Then, it holds under (A-i) that $E_{\boldsymbol{\theta}}(\mathrm{tr}(\widehat{\boldsymbol{\Sigma}_i^2})) = \mathrm{tr}(\boldsymbol{\Sigma}_i^2)$ and

$$Var_{\boldsymbol{\theta}}\left(\frac{\mathrm{tr}(\widehat{\boldsymbol{\Sigma}_i^2})}{\mathrm{tr}(\boldsymbol{\Sigma}_i^2)}\right) = \frac{4}{m^2}(1+o(1)) + \frac{8\mathrm{tr}(\boldsymbol{\Sigma}_i^4)}{\mathrm{tr}(\boldsymbol{\Sigma}_i^2)^2 m}(1+o(1))$$

as $p \to \infty$ and $m \to \infty$. One might consider $\mathrm{tr}(\widehat{\boldsymbol{\Sigma}_i^2})$ for $\mathrm{tr}(\boldsymbol{S}_{im(1)}\boldsymbol{S}_{im(2)})$ in (8). It should be noted that $\mathrm{tr}(\widehat{\boldsymbol{\Sigma}_i^2})$ is not unbiased unless (A-i) holds. In addition, it does not hold that $Var_{\boldsymbol{\theta}}(\mathrm{tr}(\widehat{\boldsymbol{\Sigma}_i^2})/\mathrm{tr}(\boldsymbol{\Sigma}_i^2)) < \infty$ when the eighth moments of each variable in $\boldsymbol{Z}_i$ are not uniformly bounded. The unbiased estimator given by Yata is quite robust in a non-Gaussian situation when compared to the estimator by Bai and Saranadasa (1996) and Srivastava (2005).

2.3 Simulation

In order to study the performance of the two-stage procedure given by (6)-(8), we took resort to computer simulations. We set $k = 2$, $p = 1600$, $b_1 = b_2 = 1$ and $W = 40$. Our goal was to construct a 95% upper-bounded confidence region, $\boldsymbol{R}_{\mathbf{N},W}$. In other words, we set $\alpha = 0.05$. Independent pseudorandom normal observations were generated for π_i : $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $i = 1, 2$. We considered the covariance matrix such as $\boldsymbol{\Sigma}_1 = c_1 \boldsymbol{B}(\rho_1^{|i-j|^{1/3}})\boldsymbol{B}$ and $\boldsymbol{\Sigma}_2 = c_2 \boldsymbol{B}(\rho_2^{|i-j|^{1/3}})\boldsymbol{B}$, where $c_i > 0$ and $\rho_i \in (0,1)$, $i = 1, 2$, and

$$\boldsymbol{B} = \mathrm{diag}\left(\sqrt{0.5 + 1/(p+1)}, \sqrt{0.5 + 2/(p+1)}, ..., \sqrt{0.5 + p/(p+1)}\right). \quad (9)$$

We considered the following three cases: (i) $(c_1, c_2) = (1, 1)$ and $(\rho_1, \rho_2) = (0.3, 0.3)$, i.e., $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$; (ii) $(c_1, c_2) = (1, 1)$ and $(\rho_1, \rho_2) = (0.3, 0.4)$, i.e., $\mathrm{tr}(\boldsymbol{\Sigma}_1) = \mathrm{tr}(\boldsymbol{\Sigma}_2)$ and $\mathrm{tr}(\boldsymbol{\Sigma}_1^2) \neq \mathrm{tr}(\boldsymbol{\Sigma}_2^2)$; (iii) $(c_1, c_2) = (1, 1.5)$ and $(\rho_1, \rho_2) = (0.3, 0.3)$, i.e., $\boldsymbol{\Sigma}_2 = 1.5 \boldsymbol{\Sigma}_1$.

Table 1 gives the findings obtained by averaging the outcomes from 2000 ($= R$, say) replications. We set $m = 20$ for each case. The findings for case (i) were given in the first block and the ones for cases (ii) and (iii) followed after the block. Under a fixed scenario, suppose that the rth replication ends with $N_i = n_{ir}$ ($i = 1, 2$) observations and the corresponding confidence region with $\boldsymbol{n}_r = (n_{1r}, n_{2r})$ for $r = 1, ..., R$. Let $\overline{n}_i = R^{-1}\sum_{r=1}^{R} n_{ir}$ and $Var(n_i) = (R-1)^{-1}\sum_{r=1}^{R}(n_{ir} - \overline{n}_i)^2$. Then, $\overline{n}$ ($= \overline{n}_1 + \overline{n}_2$) estimates $C = C_1 + C_2$ defined by (5) with its estimated variance $Var(n)$, computed analogously. 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 (ii). We had $C_1 = 86.93$, $C_2 = 86.93$ and $C = 173.86$ from (5). From 2000

Table 1 Required sample size and the coverage probability by (6)-(8) with $W = 40$

	$\overline{n}$	$\overline{n} - C$	$Var(n)$	$\overline{P}$	$s(\overline{P})$
Case (i): $(c_1, c_2) = (1, 1)$ and $(\rho_1, \rho_2) = (0.3, 0.3)$					
C	172.20	173.10	0.90	2.44	0.946	0.00505
C_1	86.10	86.54	0.44	0.79		
C_2	86.10	86.56	0.46	0.73		
Case (ii): $(c_1, c_2) = (1, 1)$ and $(\rho_1, \rho_2) = (0.3, 0.4)$					
C	173.86	174.82	0.95	3.77	0.953	0.00473
C_1	86.93	87.42	0.49	0.95		
C_2	86.93	87.40	0.47	1.31		
Case (iii): $(c_1, c_2) = (1, 1.5)$ and $(\rho_1, \rho_2) = (0.3, 0.3)$					
C	213.07	214.02	0.94	3.84	0.949	0.00491
C_1	95.77	96.26	0.48	0.96		
C_2	117.30	117.76	0.46	1.38		

independent replications, we observed $\overline{n}_1 = 87.42$ ($\overline{n}_1 - C_1 = 0.49$), $\overline{n}_2 = 87.40$ ($\overline{n}_2 - C_2 = 0.47$), $\overline{n} = 174.82$ ($\overline{n} - C = 0.95$) and $\overline{p} = 0.953$ together with $Var(n_1) = 0.95$, $Var(n_2) = 1.31$, $Var(n) = 3.77$ and $s(\overline{p}) = 0.00473$. Throughout, we observed that the required confidence regions were successfully constructed.

3 Given-Width Confidence Interval for Squared Norm of Mean Vectors

In this section, we consider constructing a confidence interval for $||\boldsymbol{\mu}||^2$ defined by

$$\boldsymbol{R}_{\mathbf{n}, \delta} = \{\boldsymbol{\mu} \in R^p : \max\{-\delta + \widetilde{T}_{\mathbf{n}}, 0\} \leq ||\boldsymbol{\mu}||^2 \leq \max\{\delta + \widetilde{T}_{\mathbf{n}}, 0\}\} \qquad (10)$$

for given $\delta \, (> 0)$. We consider an estimator of $||\boldsymbol{\mu}||^2$ as $\widetilde{T}_{\mathbf{n}} = ||\boldsymbol{T}_{\mathbf{n}}||^2 - \widehat{\Sigma}_{\mathbf{n}}$, where $\widehat{\Sigma}_{\mathbf{n}} = \sum_{i=1}^{k} b_i^2 \mathrm{tr}(\boldsymbol{S}_{in_i})/n_i$ with $\boldsymbol{S}_{in_i} = \sum_{j=1}^{n_i} (\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_{in_i})(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_{in_i})^T/(n_i - 1)$. It is shown that $E_{\boldsymbol{\theta}}(\widetilde{T}_{\mathbf{n}}) = ||\boldsymbol{\mu}||^2$ and

$$Var_{\boldsymbol{\theta}}(\widetilde{T}_{\mathbf{n}}) = 2 \sum_{i=1}^{k} \frac{b_i^4 \mathrm{tr}(\boldsymbol{\Sigma}_i^2)}{n_i(n_i - 1)} + 4 \sum_{i>j} \frac{b_i^2 b_j^2 \mathrm{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j)}{n_i n_j} + 4 \sum_{i=1}^{k} \frac{b_i^2 \boldsymbol{\mu}^T \boldsymbol{\Sigma}_i \boldsymbol{\mu}}{n_i}$$

$$= 2 \sum_{i,j} \frac{b_i^2 b_{i'}^2 \mathrm{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_{i'})}{n_i n_{i'}} + 4 \sum_{i=1}^{k} \frac{b_i^2 \boldsymbol{\mu}^T \boldsymbol{\Sigma}_i \boldsymbol{\mu}}{n_i} + O\Big(\sum_{i=1}^{k} \frac{\mathrm{tr}(\boldsymbol{\Sigma}_i^2)}{n_i^3}\Big).$$

Our goal is to construct a confidence region $\boldsymbol{R}_{\mathbf{n}, \delta}$ such that

$$P_{\boldsymbol{\theta}}(\boldsymbol{\mu} \in \boldsymbol{R}_{\mathbf{n}, \delta}) \geq 1 - \alpha \qquad (11)$$

for given $\alpha \in (0, 1)$. We assume that $\delta/\min_{1 \leq i \leq k} \mathrm{tr}(\boldsymbol{\Sigma}_i^2)^{1/2} = o(1)$.

3.1 Aysmptotic Normality and Sample Size Determination

For $\widetilde{T}_{\mathbf{n}}$, we have the following theorem.

Theorem 3.1. *Assume that $\boldsymbol{\mu}^T \boldsymbol{\Sigma}_i \boldsymbol{\mu}/(tr(\boldsymbol{\Sigma}_i^2)/n_i) = o(1)$, $i = 1, ..., k$. Assume either (A-ii) or (A-iii) with (A-v). Then, it holds as $p \to \infty$ that*

$$\frac{\widetilde{T}_{\mathbf{n}} - ||\boldsymbol{\mu}||^2}{\sqrt{Var_{\boldsymbol{\theta}}(\widetilde{T}_{\mathbf{n}})}} \Rightarrow N(0, 1).$$

From Theorem 3.1, we consider n_i's such that

$$\min\left((\delta/p^{1/2}) \sum_{i=1}^{k} n_i\right) \quad \text{subject to} \quad \frac{\delta}{\sqrt{Var_{\boldsymbol{\theta}}(\widetilde{T}_{\mathbf{n}})}} \geq z_{\alpha/2}$$

as $p \to \infty$. Throughout this section, we assume the following extra assumption:

(A-vi) $\dfrac{\boldsymbol{\mu}^T \boldsymbol{\Sigma}_i \boldsymbol{\mu}}{\delta \sqrt{\text{tr}(\boldsymbol{\Sigma}_i^2)}} = o(1)$ for $i = 1, ..., k$.

Then, we find the sample size for each π_i as

$$n_i \geq \frac{z_{\alpha/2}\sqrt{2}}{\delta} |b_i| \text{tr}(\boldsymbol{\Sigma}_i^2)^{1/4} \sum_{j=1}^{k} |b_j| \text{tr}(\boldsymbol{\Sigma}_j^2)^{1/4}$$

$$+ \frac{2z_{\alpha/2}^2}{\delta^2} |b_i| \text{tr}(\boldsymbol{\Sigma}_i^2)^{1/4} \sum_{j=1}^{k} |b_j| \frac{\boldsymbol{\mu}^T \boldsymbol{\Sigma}_j \boldsymbol{\mu}}{\text{tr}(\boldsymbol{\Sigma}_j^2)^{1/4}} \quad (= C_i, \text{ say}). \qquad (12)$$

Note that $C_i = O(p^{1/2}/\delta) = o(p)$ under (A-iv). Thus it holds that $C_i/p \to 0$ as $p \to \infty$, $i = 1, ..., k$. Then, we have the following theorem.

Theorem 3.2. *Assume either (A-ii) or (A-iii) with (A-v). For n_i satisfying (12), we have as $p \to \infty$ that*

$$\liminf P_{\boldsymbol{\theta}}(\boldsymbol{\mu} \in \boldsymbol{R}_{\mathbf{n}, \delta}) \geq 1 - \alpha.$$

3.2 Two-Stage Procedure

Since $\boldsymbol{\Sigma}_i$'s are unknown, it is necessary to estimate C_i's in (12) with some pilot samples. We propose a two-stage estimation procedure to determine the sample sizes $\boldsymbol{n}$. We suppose the following assumption: There exists a known and positive lower bound $\sigma_{i\star}$ for $\sqrt{\text{tr}(\boldsymbol{\Sigma}_i^2)}$ such that $\sigma_{i\star}/\sqrt{\text{tr}(\boldsymbol{\Sigma}_i^2)} \in (0, 1)$, $i = 1, ..., k$, as $p \to \infty$. We proceed the following two steps:

1. Let $\tau_\star = \min_{1\le i\le k} |b_i|\sqrt{\sigma_{i\star}} \sum_{j=1}^k |b_j|\sqrt{\sigma_{j\star}}$. Having a fixed integer m_0 (≥ 4), define

$$m = \max\left\{ m_0,\ \left[\frac{z_{\alpha/2}\sqrt{2}}{\delta}\tau_\star\right] + 1 \right\}. \tag{13}$$

According to (13), take pilot samples $\boldsymbol{x}_{ij}$, $j = 1,...,m$, of size m from each π_i. Then, calculate $\overline{\boldsymbol{x}}_{im(1)}$, $\overline{\boldsymbol{x}}_{im(2)}$, $\boldsymbol{S}_{im}$, $\boldsymbol{S}_{im(1)}$ and $\boldsymbol{S}_{im(2)}$ according to (7) for each π_i. We consider an unbiased estimator of $\boldsymbol{\mu}^T\boldsymbol{\Sigma}_i\boldsymbol{\mu}$ such as

$$U_i = \Big((b_i\overline{\boldsymbol{x}}_{im(1)} + \boldsymbol{T}_{m(i)})^T \boldsymbol{S}_{im(2)}(b_i\overline{\boldsymbol{x}}_{im(1)} + \boldsymbol{T}_{m(i)})$$

$$+ (b_i\overline{\boldsymbol{x}}_{im(2)} + \boldsymbol{T}_{m(i)})^T \boldsymbol{S}_{im(1)}(b_i\overline{\boldsymbol{x}}_{im(2)} + \boldsymbol{T}_{m(i)}) \Big)/2 - \frac{b_i^2\mathrm{tr}(\boldsymbol{S}_{im(1)}\boldsymbol{S}_{im(2)})m}{2m_1 m_2}$$

$$- \sum_{j(\ne i)}^k \frac{b_j^2\mathrm{tr}((\boldsymbol{S}_{im(1)} + \boldsymbol{S}_{im(2)})\boldsymbol{S}_{jm})}{2m} \tag{14}$$

with $\boldsymbol{T}_{m(i)} = \sum_{j(\ne i)}^k b_j\overline{\boldsymbol{x}}_{jm}$. Note that $E_{\boldsymbol{\theta}}(U_i) = \boldsymbol{\mu}^T\boldsymbol{\Sigma}_i\boldsymbol{\mu}$. Define the total sample size for each π_i by

$$N_i = \max\left\{ m,\ \left[\frac{z_{\alpha/2}\sqrt{2}}{\delta}|b_i|\mathrm{tr}(\boldsymbol{S}_{im(1)}\boldsymbol{S}_{im(2)})^{1/4} \sum_{j=1}^k |b_j|\mathrm{tr}(\boldsymbol{S}_{jm(1)}\boldsymbol{S}_{jm(2)})^{1/4} \right.\right.$$

$$\left.\left. + \frac{2z_{\alpha/2}^2}{\delta^2}|b_i|\mathrm{tr}(\boldsymbol{S}_{im(1)}\boldsymbol{S}_{im(2)})^{1/4} \sum_{j=1}^k |b_j|\frac{\max\{U_j,\ 0\}}{\mathrm{tr}(\boldsymbol{S}_{jm(1)}\boldsymbol{S}_{jm(2)})^{1/4}} \right] + 1 \right\}. \tag{15}$$

Let $\boldsymbol{N} = (N_1,...,N_k)$.

2. Take additional samples $\boldsymbol{x}_{ij}$, $j = m+1,...,N_i$, of size $N_i - m$ from each π_i. By combining the initial samples and the additional samples, calculate $\widetilde{T}_{\boldsymbol{N}} = \|\boldsymbol{T}_{\boldsymbol{N}}\|^2 - \widehat{\Sigma}_{\boldsymbol{N}}$, where $\widehat{\Sigma}_{\boldsymbol{N}} = \sum_{i=1}^k b_i^2\mathrm{tr}(\boldsymbol{S}_{iN_i})/N_i$ with $\boldsymbol{S}_{iN_i} = (N_i - 1)^{-1}\sum_{j=1}^{N_i}(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_{iN_i})(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_{iN_i})^T$. Then, define $\boldsymbol{R}_{\boldsymbol{N},\delta}$ according to (10).

We have the following theorem.

Theorem 3.3. *Assume either (A-ii) or (A-iii) with (A-v). Assume also that $\max_{1\le i\le k}\{tr(\boldsymbol{\Sigma}_i^4)\} = O(p\delta^2)$. For the two-stage procedure given by (13)-(15), we have as $p \to \infty$ that*

$$\liminf P_{\boldsymbol{\theta}}(\boldsymbol{\mu} \in \boldsymbol{R}_{\boldsymbol{N},\delta}) \ge 1 - \alpha.$$

Remark 6. Assume either (A-ii) or (A-iii). It holds as $p \to \infty$ that $N_i/C_i = 1 + o_p(1)$, so that $N_i/p = o_p(1)$ for $i = 1,...,k$. It should be noted that the result given by Theorem 3.3 can be claimed in a HDLSS situation.

Remark 7. One of the choices of $\sigma_{i\star}$ is, for example, a positive lower bound, σ_{i0}, for $\mathrm{tr}(\boldsymbol{\Sigma}_i)/\sqrt{p}$ such that $\sigma_{i0}\sqrt{p}/\mathrm{tr}(\boldsymbol{\Sigma}_i) \in (0,1)$ as $p \to \infty$. Then, it holds from Schwartz's inequality and (A-iv) that

$$0 < \sigma_{i0}/\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_i^2)} = (\sigma_{i0}\sqrt{p}/\mathrm{tr}(\boldsymbol{\Sigma}_i))(\mathrm{tr}(\boldsymbol{\Sigma}_i)/\sqrt{p\,\mathrm{tr}(\boldsymbol{\Sigma}_i^2)}) < 1$$

as $p \to \infty$. We emphasize that the result in Theorem 3.3 can be claimed as long as $\sigma_{i\star}/p^{1/2} > 0$ as $p \to \infty$ for $i = 1, ..., k$. In that sense, the two-stage procedure is quite robust for the misidentification of $\sigma_{i\star}$.

Remark 8. For (14), under (A-iv), (A-vi) and either (A-ii) or (A-iii), it holds as $p \to \infty$ that

$$Var_{\boldsymbol{\theta}}(U_i) = o(p^2/m^2);$$
$$U_i = \boldsymbol{\mu}^T \boldsymbol{\Sigma}_i \boldsymbol{\mu} + o_p(p/m) \quad \text{for} \quad i = 1, ..., k.$$

Remark 9. Under (A-i), one may consider

$$U_i = \boldsymbol{T}_m^T \boldsymbol{S}_{im} \boldsymbol{T}_m - \frac{b_i^2 \mathrm{tr}(\widehat{\boldsymbol{\Sigma}_i^2})}{m} - \sum_{j(\neq i)}^k \frac{b_j^2 \mathrm{tr}(\boldsymbol{S}_{im} \boldsymbol{S}_{jm})}{m}$$

as an unbiased estimator instead of (14), where $\boldsymbol{T}_m = \sum_{i=1}^k b_i \overline{\boldsymbol{x}}_{im}$ and $\mathrm{tr}(\widehat{\boldsymbol{\Sigma}_i^2})$ is given in Remark 5. Then, it holds that $E_{\boldsymbol{\theta}}(U_i) = \boldsymbol{\mu}^T \boldsymbol{\Sigma}_i \boldsymbol{\mu}$.

3.3 Simulation

In order to study the performance of the two-stage procedure given by (13)-(15), we took resort to computer simulations. We set $k = 2$, $b_1 = 1$, $b_2 = -1$ and $\delta = 10$. Our goal was to construct a 95% confidence interval $\boldsymbol{R}_{\mathbf{N}, \delta}$. In other words, we set $\alpha = 0.05$. We considered a non-Gaussian case such as $z_{ijl} = (13/15)^{1/2} w_{ijl}$, where w_{ijl}, $i = 1, 2$; $j = 1, ..., p$ $(l = 1, ..., n_i)$ are independently generated by t-distribution with 15 degrees of freedom. Then, note that $E(z_{ijl}) = 0$, $E(z_{ijl}^2) = 1$ and z_{ijl}, $j = 1, ..., p$ $(i = 1, 2; l = 1, ..., n_i)$ are independent. Let $\boldsymbol{x}_{il} = \boldsymbol{H} \boldsymbol{\Lambda}_i^{1/2} (z_{i1l}, ..., z_{ipl})^T + \boldsymbol{\mu}_i$ $(i = 1, 2; l = 1, ..., n_i)$ with $\boldsymbol{\mu}_1 = (1, ..., 1, 0, ..., 0)^T$ whose first 25 elements are 1, $\boldsymbol{\mu}_2 = (0, ..., 0)^T$, $\boldsymbol{\Sigma}_1 = \boldsymbol{H} \boldsymbol{\Lambda}_1 \boldsymbol{H}^T = \boldsymbol{B}(0.3^{|i-j|^{1/3}})\boldsymbol{B}$ and $\boldsymbol{\Sigma}_2 = \boldsymbol{H} \boldsymbol{\Lambda}_2 \boldsymbol{H}^T = 1.2\boldsymbol{B}(0.3^{|i-j|^{1/3}})\boldsymbol{B}$, where $\boldsymbol{B}$ is given by (9). Then, the population distributions of $\boldsymbol{x}_{il}$, $i = 1, 2$, satisfy (A-ii). Note that $||\boldsymbol{\mu}||^2 = 25$. We set $\sigma_{i\star} = 0.8 \times \sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_i^2)}$, $i = 1, 2$. Then, we obtained $m = 13, 18, 22, 25, 28$ from (13) for $p = 400(400)2000$, respectively.

In table 2, each block gives the findings when $p = 400(400)2000$. The findings were obtained by averaging the outcomes from $2000(= R$, say) replications. Under a fixed scenario, suppose that the rth replication ends with $N_i = n_{ir}$ $(i = 1, 2)$ observations and the corresponding confidence interval with $\boldsymbol{n}_r = (n_{1r}, n_{2r})$ for $r = 1, ..., R$. Let $\overline{n}_i = R^{-1} \sum_{r=1}^R n_{ir}$ and $Var(n_i/C_i) = (R-1)^{-1} \sum_{r=1}^R (n_{ir} - \overline{n}_i)^2 / C_i^2$. Then, $\overline{n}$ $(= \overline{n}_1 + \overline{n}_2)$ estimates $C = C_1 + C_2$ defined by (12) with its estimated variance, $Var(n/C)$, computed analogously. In the end of the rth replication, we checked whether $\boldsymbol{\mu}$ does (or does not) belong to the corresponding confidence interval and defined $P_r = 1$ (or 0)

Table 2 Required sample size and the coverage probability by (13)-(15) with $\delta = 10$

	$\overline{n}$	$\overline{n} - C$	$Var(n/C)$	$\overline{P}$	$s(\overline{P})$
When $p = 400$: $m = 13$					
C	48.94	56.30	7.36	0.195	0.939	0.00537
C_1	23.36	26.89	3.53	0.199		
C_2	25.59	29.41	3.83	0.199		
When $p = 800$: $m = 18$					
C	61.69	70.57	8.89	0.132	0.935	0.00553
C_1	29.44	33.74	4.30	0.135		
C_2	32.25	36.84	4.59	0.133		
When $p = 1200$: $m = 22$					
C	71.68	81.68	10.0	0.112	0.943	0.00521
C_1	34.21	39.04	4.83	0.114		
C_2	37.47	42.64	5.16	0.113		
When $p = 1600$: $m = 25$					
C	80.16	91.49	11.33	0.0971	0.946	0.00505
C_1	38.25	43.69	5.44	0.0986		
C_2	41.91	47.80	5.89	0.0973		
When $p = 2000$: $m = 28$					
C	87.65	98.56	10.90	0.0764	0.950	0.00487
C_1	41.83	47.07	5.24	0.0775		
C_2	45.82	51.48	5.66	0.0765		

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 $s^2(\overline{P}) = R^{-1}\overline{P}(1 - \overline{P})$. Throughout, we observed that the two-stage procedure constructed required confidence intervals successfully for HDLSS situations.

4 Data Analysis

In this section, we demonstrate how the new methodologies perform for a real data set in HDLSS data situations. We analyzed gene expression data given by Chiaretti et al. (2004) in which data set consisted of 12625 ($= p$) genes and 128 samples. Note that the expression measures were obtained 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, π_1: B-cell and π_2: T-cell. We set $\boldsymbol{\mu} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ ($b_1 = 1$, $b_2 = -1$). We assumed (A-iv) and either (A-ii) or (A-iii) with (A-v).

We first considered constructing a confidence region for $\boldsymbol{\mu}$ along the lines of Section 2. We set $\alpha = 0.05$ and $W = 300$. We assumed that $\text{tr}(\boldsymbol{\Sigma}_1) > 1500$ for B-cell and $\text{tr}(\boldsymbol{\Sigma}_2) > 1500$ for T-cell. We set $\sigma_{1\star} = 1500$ and $\sigma_{2\star} = 1500$ so that $\tau_\star = \min_{i=1,2} \sigma_{i\star}^{1/2}(\sigma_{1\star}^{1/2} + \sigma_{2\star}^{1/2}) = 3000$. We calculated the pilot sample

size as

$$m = \max \left\{ 4, \left[\frac{3000}{300} \right] + 1 \right\} = 10$$

according to (6). So, we took the first 10 samples from each π_i as a pilot sample. We calculated $\text{tr}(\boldsymbol{S}_{1m}) = 2646$, $\text{tr}(\boldsymbol{S}_{2m}) = 2664$, $\text{tr}(\boldsymbol{S}_{1m(1)}\boldsymbol{S}_{1m(2)}) = 7.22 \times 10^5$ and $\text{tr}(\boldsymbol{S}_{2m(1)}\boldsymbol{S}_{2m(2)}) = 1.09 \times 10^5$ from (7). Note that $z_\alpha = 1.64$. Then, from (8), we had the total sample sizes as

$$N_1 = \max \left\{ 10, \left[\frac{\sqrt{2646}}{300} \left(\sqrt{2646} + \sqrt{2664} \right) \right. \right.$$
$$\left. \left. + \frac{1.64\sqrt{2}}{300} \sqrt{2646} \left(\sqrt{\frac{7.22 \times 10^5}{2646}} + \sqrt{\frac{1.09 \times 10^5}{2664}} \right) \right] + 1 \right\} = 27,$$

$$N_2 = 27.$$

So, we took the next 17 samples from π_1 and the next 17 samples from π_2. Then, we had an estimate of $\boldsymbol{\mu}$ by

$$\boldsymbol{T_N} = \overline{\boldsymbol{x}}_{1N_1} - \overline{\boldsymbol{x}}_{2N_2} = (-0.102, -0.014, 0.026, ..., 0.095, 0.114, 0.174)^T,$$

so that a 95% confidence region for $\boldsymbol{\mu}$ by

$$\boldsymbol{R_{N,W}} = \{ \boldsymbol{\mu} \in R^p : ||\boldsymbol{T_N} - \boldsymbol{\mu}||^2 \le 300 \}. \tag{16}$$

We checked whether $\boldsymbol{\mu} = \boldsymbol{0}$ or not. When $\boldsymbol{\mu} = \boldsymbol{0}$, it followed that $||\boldsymbol{T_N} - \boldsymbol{\mu}||^2 = ||\boldsymbol{T_N}||^2 = 1746$ (> 300). Since $\boldsymbol{0} \notin \boldsymbol{R_{N,W}}$, we concluded that $\boldsymbol{\mu} \ne \boldsymbol{0}$. Let $\boldsymbol{T_N} = (T_{1\mathbf{N}}, ..., T_{p\mathbf{N}})^T$. We considered a variable selection procedure by

$$T_{j\mathbf{N}(*)} = \begin{cases} T_{j\mathbf{N}} & \text{if } |T_{j\mathbf{N}}| \ge 0.4, \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

for $j = 1, ..., p$. Let $\boldsymbol{T_{N(*)}} = (T_{1\mathbf{N}(*)}, ..., T_{p\mathbf{N}(*)})^T$. Then, we had that

$$\boldsymbol{T_{N(*)}} = (0, 0, 0, 0, 0, 0.572, ..., 0, 0, 0)^T.$$

The number of nonzero elements in $\boldsymbol{T_{N(*)}}$ was 1709. In other words, the selection procedure (17) chose 1709 genes from 12625 genes. When $\boldsymbol{\mu} = \boldsymbol{T_{N(*)}}$, it followed that $||\boldsymbol{T_N} - \boldsymbol{\mu}||^2 = ||\boldsymbol{T_N} - \boldsymbol{T_{N(*)}}||^2 = 263$ (< 300). Since $\boldsymbol{\mu} = \boldsymbol{T_{N(*)}} \in \boldsymbol{R_{N,W}}$, we considered $\boldsymbol{T_{N(*)}}$ as an estimate of $\boldsymbol{\mu}$.

Next, we considered constructing a confidence interval for $||\boldsymbol{\mu}||^2$ along the lines of Section 3. We set $\alpha = 0.05$ and $\delta = 200$. We assumed that $\text{tr}(\boldsymbol{\Sigma}_1^2)^{1/2} > 600$ for B-cell and $\text{tr}(\boldsymbol{\Sigma}_2^2)^{1/2} > 300$ for T-cell. We set $\sigma_{1\star} = 600$ and $\sigma_{2\star} = 300$ so that $\tau_\star = \min_{i=1,2} \sigma_{i\star}^{1/2}(\sigma_{1\star}^{1/2} + \sigma_{2\star}^{1/2}) = 724$. Note that $z_{\alpha/2} = 1.96$. We calculated the pilot sample size as

$$m = \max \left\{ 4, \left[\frac{1.96\sqrt{2} \times 724}{200} \right] + 1 \right\} = 11$$

according to (13). So, we took the first 11 samples from each π_i as a pilot sample. We calculated $\text{tr}(\boldsymbol{S}_{1m(1)}\boldsymbol{S}_{1m(2)}) = 6.28 \times 10^5$ and $\text{tr}(\boldsymbol{S}_{2m(1)}\boldsymbol{S}_{2m(2)}) = 3.63 \times 10^5$ from (7) so that $\max(U_1, \ 0) = 0$ and $\max(U_2, \ 0) = 60400$ from (14). Then, from (15), we had the total sample sizes as

$$N_1 = \max\left\{11, \ \left[\frac{1.96\sqrt{2}}{200}(6.28 \times 10^5)^{1/4}\left\{(6.28 \times 10^5)^{1/4} + (3.63 \times 10^5)^{1/4}\right\}\right.\right.$$
$$\left.\left. + \frac{1.96^2 2}{200^2}(6.28 \times 10^5)^{1/4}\left(0 + \frac{60400}{(3.63 \times 10^5)^{1/4}}\right)\right] + 1\right\} = 34,$$

$$N_2 = 30.$$

So, we took the next 23 samples from π_1 and the next 19 samples from π_2. Then, we had $||\boldsymbol{T_N}||^2 = 1692$ and $\widehat{\Sigma}_{\mathbf{N}} = \text{tr}(\boldsymbol{S}_{1N_1})/N_1 + \text{tr}(\boldsymbol{\Sigma}_{2N_2})/N_2 = 149$ so that $\widetilde{T}_{\mathbf{N}} = ||\boldsymbol{T_N}||^2 - \widehat{\Sigma}_{\mathbf{N}} = 1543$. Finally, we gave a 95% confidence interval by

$$\boldsymbol{R}_{\mathbf{N},\delta} = \{\boldsymbol{\mu} \in R^p : \max\{-\delta + \widetilde{T}_{\mathbf{N}}, 0\} \leq ||\boldsymbol{\mu}||^2 \leq \{\delta + \widetilde{T}_{\mathbf{N}}, 0\}\}$$
$$= \{\boldsymbol{\mu} \in R^p : 1343 \leq ||\boldsymbol{\mu}||^2 \leq 1743\}.$$

We emphasize that one may apply a confidence interval for $||\boldsymbol{\mu}||^2$ to the discriminant analysis for HDLSS data. (Refer to Section 4 in Aoshima and Yata, 2011).

A Appendix

Proof of Theorem 2.2. We have from (5) that $\Sigma_{\mathbf{n}} \leq W/(1 + u(\alpha))$, where

$$u(\alpha) = z_\alpha\sqrt{2}\sum_{i=1}^{k}|b_i|\sqrt{\frac{\text{tr}(\boldsymbol{\Sigma}_i^2)}{\text{tr}(\boldsymbol{\Sigma}_i)}}\Big/\left(\sum_{j=1}^{k}|b_j|\sqrt{\text{tr}(\boldsymbol{\Sigma}_j)}\right). \tag{18}$$

It holds that

$$\sqrt{2\sum_{i,j}b_i^2 b_j^2 \text{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j)/(n_i n_j)} \leq \sqrt{2}\sum_{i=1}^{k}b_i^2\sqrt{\text{tr}(\boldsymbol{\Sigma}_i^2)}/n_i \leq \frac{u(\alpha)W}{z_\alpha(1 + u(\alpha))}.$$

Then, we have from (4) that

$$P_{\boldsymbol{\theta}}(||\boldsymbol{T_n} - \boldsymbol{\mu}||^2 \leq W) \geq P_{\boldsymbol{\theta}}\left(||\boldsymbol{T_n} - \boldsymbol{\mu}||^2 - \Sigma_{\mathbf{n}} \leq \frac{u(\alpha)W}{1 + u(\alpha)}\right)$$
$$\geq P_{\boldsymbol{\theta}}(N(0,1) \leq z_\alpha) + o(1) = 1 - \alpha + o(1).$$

It concludes the result. □

Proof of Theorem 2.3. We have under (A-iv) and either (A-ii) or (A-iii) that

$$Var_{\boldsymbol{\theta}}\left(\frac{\text{tr}(\boldsymbol{S}_{im})}{\text{tr}(\boldsymbol{\Sigma}_i)}\right) = O(\text{tr}(\boldsymbol{\Sigma}_i^2)/(\text{tr}(\boldsymbol{\Sigma}_i)^2 m)) = O((mp)^{-1}),$$

$$Var_{\boldsymbol{\theta}}\left(\frac{\text{tr}(\boldsymbol{S}_{im(1)}\boldsymbol{S}_{im(2)})}{\text{tr}(\boldsymbol{\Sigma}_i^2)}\right) = O(m^{-2}) + O(\text{tr}(\boldsymbol{\Sigma}_i^4)/(\text{tr}(\boldsymbol{\Sigma}_i^2)^2 m)) = o(m^{-1}). \tag{19}$$

Then, it holds as $p \to \infty$ that $|N_i - C_i| = O_p(1)$ under (A-iv) and either (A-ii) or (A-iii). Let $C_{i\star} = [C_i - m^\xi]$ $(i = 1, ..., k)$ for $\xi \in (0, 1/2)$. We have as $p \to \infty$ that $m < C_{i\star} < N_i < C_i + m^\xi$ w.p.1. Here, we write that $\overline{\boldsymbol{x}}_{iN_i} = \sum_{l=1}^{C_{i\star}} \boldsymbol{x}_{il}/N_i + \sum_{l=C_{i\star}+1}^{N_i} \boldsymbol{x}_{il}/N_i$. Then, it holds that

$$
||\overline{\boldsymbol{x}}_{iN_i} - \boldsymbol{\mu}_i||^2 - \mathrm{tr}(\boldsymbol{S}_{iN_i})/N_i
$$
$$
= \sum_{j=1}^{p} \lambda_{ij} \left(\frac{\sum_{l \neq l'}^{C_{i\star}} z_{ijl}z_{ijl'}}{N_i(N_i-1)} + 2\frac{\sum_{l=1}^{C_{i\star}}\sum_{l'=C_{i\star}+1}^{N_i} z_{ijl}z_{ijl'}}{N_i(N_i-1)} + \frac{\sum_{l \neq l'(\geq C_{i\star}+1)}^{N_i} z_{ijl}z_{ijl'}}{N_i(N_i-1)} \right).
$$

By using Chebyshev's inequality and Schwarz's inequality, for any $\tau > 0$, we have that

$$
P_{\boldsymbol{\theta}} \left(\sum_{l \neq l'(\geq C_{i\star}+1)}^{N_i} \left| \sum_{j=1}^{p} \frac{\lambda_{ij}z_{ijl}z_{ijl'}}{C_{i\star}^2} \right| > \tau W/p^{1/2} \right)
$$
$$
\leq P_{\boldsymbol{\theta}} \left(\sum_{l \neq l'(\geq C_{i\star}+1)}^{C_i+[m^\xi]+1} \left| \sum_{j=1}^{p} \frac{\lambda_{ij}z_{ijl}z_{ijl'}}{C_{i\star}^2} \right| > \tau W/p^{1/2} \right) + o(1) = O\left(\frac{m^{4\xi}}{C_{i\star}^2} \right) + o(1) = o(1).
$$
$$(20)$$

Thus we claim that

$$
\sum_{j=1}^{p} \frac{\lambda_{ij} \sum_{l \neq l'(\geq C_{i\star}+1)}^{N_i} z_{ijl}z_{ijl'}}{N_i(N_i-1)} = o_p(W/p^{1/2}).
$$

Similarly to (20), for any $\tau > 0$, we have that

$$
P_{\boldsymbol{\theta}} \left(\sum_{l'=C_{i\star}+1}^{N_i} \left| \frac{\sum_{j=1}^{p} \lambda_{ij} \sum_{l=1}^{C_{i\star}} z_{ijl}z_{ijl'}}{C_{i\star}^2} \right| > \tau W/p^{1/2} \right) = o(1).
$$

Thus we have that

$$
\sum_{j=1}^{p} \frac{\lambda_{ij} \sum_{l=1}^{C_{i\star}} \sum_{l'=C_{i\star}+1}^{N_i} z_{ijl}z_{ijl'}}{N_i(N_i-1)} = o_p(W/p^{1/2}).
$$

Then, it holds as $p \to \infty$ that

$$
||\overline{\boldsymbol{x}}_{iN_i} - \boldsymbol{\mu}_i||^2 - \mathrm{tr}(\boldsymbol{S}_{iN_i})/N_i = \sum_{j=1}^{p} \lambda_{ij} \left(\frac{\sum_{l \neq l'}^{C_{i\star}} z_{ijl}z_{ijl'}}{C_{i\star}(C_{i\star}-1)} \right) + o_p(W/p^{1/2}).
$$

Similarly, it holds for $i \neq i'$ that

$$
(\overline{\boldsymbol{x}}_{iN_i} - \boldsymbol{\mu}_i)^T (\overline{\boldsymbol{x}}_{i'N_{i'}} - \boldsymbol{\mu}_{i'}) = (\overline{\boldsymbol{x}}_{iC_{i\star}} - \boldsymbol{\mu}_i)^T (\overline{\boldsymbol{x}}_{i'C_{i'\star}} - \boldsymbol{\mu}_{i'}) + o_p(W/p^{1/2}).
$$

Then, we have that

$$
||\boldsymbol{T}_{\mathbf{N}} - \boldsymbol{\mu}||^2 - \widehat{\Sigma}_{\mathbf{N}} = ||\boldsymbol{T}_{\boldsymbol{C}_\star} - \boldsymbol{\mu}||^2 - \widehat{\Sigma}_{\boldsymbol{C}_\star} + o_p(W/p^{1/2}),
$$

where $\boldsymbol{C}_\star = (C_{1\star}, ..., C_{k\star})$. Here, let us write that

$$
Y_i = \frac{1}{W} |b_i| \sqrt{\mathrm{tr}(\boldsymbol{S}_{im})} \sum_{j=1}^{k} |b_j| \sqrt{\mathrm{tr}(\boldsymbol{S}_{jm})} + \frac{z_\alpha \sqrt{2}}{W} |b_i| \sqrt{\mathrm{tr}(\boldsymbol{S}_{im})} \sum_{j=1}^{k} |b_j| \sqrt{\frac{\mathrm{tr}(\boldsymbol{S}_{jm(1)}\boldsymbol{S}_{jm(2)})}{\mathrm{tr}(\boldsymbol{S}_{jm})}}.
$$

From (19), it holds that $Y_i = C_i + o_p(1)$. Then we have that $N_i \geq [Y_i]+1 = [C_i+o_p(1)]+1 \geq C_i$ w.p.1. Thus we claim that $\Sigma_{\mathbf{N}} - \Sigma_{\boldsymbol{C}} \leq 0$ w.p.1, where $\boldsymbol{C} = (C_1, ..., C_k)$. Note that

$\widehat{\Sigma}_{\mathbf{N}} - \Sigma_{\mathbf{N}} = o_p(W/p^{1/2})$, $\widehat{\Sigma}_{\boldsymbol{C}_\star} - \Sigma_{\boldsymbol{C}_\star} = o_p(W/p^{1/2})$ and $u(\alpha) = O(p^{-1/2})$, where $u(\alpha)$ is defined by (18). Therefore, similarly to the proof of Theorem 2.2, we have that

$$P_{\boldsymbol{\theta}}(||\boldsymbol{T_\mathbf{N}} - \boldsymbol{\mu}||^2 - \widehat{\Sigma}_{\mathbf{N}} \leq W - \widehat{\Sigma}_{\mathbf{N}}) \geq P_{\boldsymbol{\theta}}\Big(||\boldsymbol{T_{C_\star}} - \boldsymbol{\mu}||^2 - \Sigma_{\boldsymbol{C}_\star} \leq W - \Sigma_{\boldsymbol{C}}\Big) + o(1)$$
$$\geq 1 - \alpha + o(1).$$

It concludes the result. $\qquad\qquad\square$

Proof of Theorem 2.4. From (A-i), (A-iv) and (19), we claim $E_{\boldsymbol{\theta}}(Y_i) = C_i + o(1)$, where Y_i is defined in the proof of Theorem 2.3. In a way similar to the proof of Lemma 2 in Yata and Aoshima (2009a), we have under (A-i) that

$$E_{\boldsymbol{\theta}}\big\{(N_i - [Y_i] - 1)^t\big\} = o(m^{t-2}) \quad (t = 1, 2).$$

Note that $Var_{\boldsymbol{\theta}}(Y_i) = o(1)$. Then, it holds that

$$|E_{\boldsymbol{\theta}}(N_i - C_i)| \leq |E_{\boldsymbol{\theta}}(Y_i) - C_i| + 1 + o(1) = 1 + o(1).$$

Note that $|E_{\boldsymbol{\theta}}([Y_i]) - [C_i]| \leq 1 + o(1)$ and $E_{\boldsymbol{\theta}}\{(Y_i - C_i)^2\} = o(1)$. Then, it holds that

$$Var_{\boldsymbol{\theta}}\{N_i\} = E_{\boldsymbol{\theta}}\{([Y_i] - E_{\boldsymbol{\theta}}\{[Y_i]\})^2\} + o(1) \leq E_{\boldsymbol{\theta}}\{(|[Y_i] - [C_i]| + 1)^2\} + o(1)$$
$$\leq E_{\boldsymbol{\theta}}\{(|Y_i - C_i| + 2)^2\} + o(1) \leq E_{\boldsymbol{\theta}}\{(|Y_i - C_i| + 3)^2\} + o(1) = O(1).$$

It concludes the results. $\qquad\qquad\square$

Proof of Theorem 3.1. From the assumptions that $\boldsymbol{\mu}^T \boldsymbol{\Sigma}_i \boldsymbol{\mu} = o(\text{tr}(\boldsymbol{\Sigma}_i^2)/n_i)$, $i = 1, ..., k$, it holds as $p \to \infty$ that $Var_{\boldsymbol{\theta}}(\boldsymbol{\mu}^T \boldsymbol{T_\mathbf{n}}) = O(\sum_{i=1}^k \boldsymbol{\mu}^T \boldsymbol{\Sigma}_i \boldsymbol{\mu}/n_i)$. Thus we have that

$$\frac{\widetilde{T}_\mathbf{n} - ||\boldsymbol{\mu}||^2}{\sqrt{Var_{\boldsymbol{\theta}}(\widetilde{T}_\mathbf{n})}} = \frac{||\boldsymbol{T_\mathbf{n}} - \boldsymbol{\mu}||^2 - \widehat{\Sigma}_\mathbf{n}}{\sqrt{2 \sum_{i,j} b_i^2 b_j^2 \text{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j)/(n_i n_j)}} + o_p(1).$$

Then, by using Corollary 2.1 in Aoshima and Yata (2011), it concludes the result. $\qquad\square$

Proof of Theorem 3.2. Note that $Var_{\boldsymbol{\theta}}(\widetilde{T}_\mathbf{n}) z_{\alpha/2}^2/\delta^2 \leq 1 + o(1)$. Then, we have from Theorem 3.1 that

$$P_{\boldsymbol{\theta}}(|\widetilde{T}_\mathbf{n} - ||\boldsymbol{\mu}||^2| \leq \delta) \geq P_{\boldsymbol{\theta}}(|N(0,1)| < z_{\alpha/2}) + o(1) = 1 - \alpha + o(1).$$

By noting that $||\boldsymbol{\mu}||^2 \geq 0$, it holds that $P_{\boldsymbol{\theta}}(\boldsymbol{\mu} \in \boldsymbol{R}_{\mathbf{n},\delta}) \geq 1 - \alpha + o(1)$. It concludes the result. $\square$

Proof of Theorem 3.3. Under either (A-ii) or (A-iii), we have that

$$Var_{\boldsymbol{\theta}}(U_i) = O\Big(\sum_{j,j'} \frac{\text{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j \boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_{j'})}{m^2}\Big) + O\Big(\sum_{j,j'} \frac{\text{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j)\text{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_{j'})}{m^3}\Big) + O\Big(\frac{(\boldsymbol{\mu}^T \boldsymbol{\Sigma}_i \boldsymbol{\mu})^2}{m}\Big)$$
$$+ O\Big(\sum_{j=1}^k \frac{\boldsymbol{\mu}^T \boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j \boldsymbol{\Sigma}_i \boldsymbol{\mu}}{m}\Big) + O\Big(\sum_{j=1}^k \frac{\text{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j)\boldsymbol{\mu}^T \boldsymbol{\Sigma}_i \boldsymbol{\mu}}{m^2}\Big). \tag{21}$$

Note that $\text{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j) \leq \max_{1 \leq i \leq k} \text{tr}(\boldsymbol{\Sigma}_i^2)$, $\text{tr}(\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j \boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_{j'}) \leq \max_{1 \leq i \leq k} \text{tr}(\boldsymbol{\Sigma}_i^4)$ and $\boldsymbol{\mu}^T \boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j \boldsymbol{\Sigma}_i \boldsymbol{\mu} \leq \boldsymbol{\mu}^T \boldsymbol{\Sigma}_i \boldsymbol{\mu} \sqrt{\text{tr}(\boldsymbol{\Sigma}_i^2 \boldsymbol{\Sigma}_j^2)}$. From the assumptions that (A-iv), (A-vi) and $\max_{1 \leq i \leq k} \text{tr}(\boldsymbol{\Sigma}_i^4) = O(p\delta^2)$ as $p \to \infty$, we have that

$$Var_{\boldsymbol{\theta}}(U_i/\delta^2) = O(m). \tag{22}$$

From (19) and (22), it holds that $|N_i - C_i| = O_p(m^{1/2})$. Let $C_{i\star} = [C_i - m^{1/2+\xi}]$ $(i = 1, ..., k)$ for $\xi \in (0, 1/6)$. We have as $p \to \infty$ that $m < C_{i\star} < N_i < C_i + m^{1/2+\xi}$ w.p.1. Here, we write that $\omega_i = [|(N_i - C_{i\star})|/[m^{1/2-\xi}]] + 1$, where ω_i is an integer. Note that $\omega_i = O_p(m^{2\xi})$. Let $\boldsymbol{R}_t = \{C_{i\star} + 1 + (t-1)[m^{1/2-\xi}], ..., C_{i\star} + t[m^{1/2-\xi}]\}$ $(t = 1, .., \omega_i - 1)$ and $\boldsymbol{R}_{\omega_i} = \{C_{i\star} + 1 + (\omega_i - 1)[m^{1/2-\xi}], ..., N_i\}$. Then, we can describe $\sum_{l=C_{i\star}+1}^{N_i}$ as $\sum_{j=1}^{\omega_i} \sum_{l \in R_j}$. Thus we have that

$$\sum_{j=1}^{p} \frac{\lambda_{ij} \sum_{l=1}^{C_{i\star}} \sum_{l'=C_{i\star}+1}^{N_i} z_{ijl} z_{ijl'}}{N_i(N_i - 1)}$$

$$= \sum_{t=1}^{\omega_i - 1} \sum_{j=1}^{p} \frac{\lambda_{ij} \sum_{l=1}^{C_{i\star}} \sum_{l' \in \boldsymbol{R}_t} z_{ijl} z_{ijl'}}{N_i(N_i - 1)} + \sum_{l' \in \boldsymbol{R}_{\omega_i}} \sum_{j=1}^{p} \frac{\lambda_{ij} \sum_{l=1}^{C_{i\star}} z_{ijl} z_{ijl'}}{N_i(N_i - 1)}.$$

In a way similar to (20), for any $\tau > 0$, we have that

$$P_{\boldsymbol{\theta}}\left(\sum_{t=1}^{\omega_i - 1} \left| \sum_{j=1}^{p} \frac{\lambda_{ij} \sum_{l=1}^{C_{i\star}} \sum_{l' \in \boldsymbol{R}_t} z_{ijl} z_{ijl'}}{C_{i\star}^2} \right| > \tau\delta \right) = O(m^{4\xi} m^{1/2-\xi}/C_{i\star}) + o(1) = o(1),$$

$$P_{\boldsymbol{\theta}}\left(\sum_{l' \in \boldsymbol{R}_{\omega_i}} \left| \sum_{j=1}^{p} \frac{\lambda_{ij} \sum_{l=1}^{C_{i\star}} z_{ijl} z_{ijl'}}{C_{i\star}^2} \right| > \tau\delta \right) = O(m^{1-2\xi}/C_{i\star}) = o(1).$$

Then, we have that

$$\sum_{j=1}^{p} \frac{\lambda_{ij} \sum_{l=1}^{C_{i\star}} \sum_{l'=C_{i\star}+1}^{N_i} z_{ijl} z_{ijl'}}{N_i(N_i - 1)} = o_p(\delta).$$

Similarly, we have that

$$\sum_{j=1}^{p} \frac{\lambda_{ij} \sum_{l \neq l' (\geq C_{i\star}+1)}^{N_i} z_{ijl} z_{ijl'}}{N_i(N_i - 1)} = o_p(\delta) \text{ and } \sum_{l=C_{i\star}+1}^{N_i} \boldsymbol{\mu}^T(\boldsymbol{x}_{il} - \boldsymbol{\mu}_i)/N_i = o_p(\delta).$$

In a way similar to the proof of Theorem 2.3, we claim that

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

where $\boldsymbol{C}_\star = (C_{1\star}, ..., C_{k\star})$. Hence, similarly to the proof of Theorem 3.2, it concludes the result. $\qquad\square$

References

1. Ahn J, Marron JS, Muller KM, Chi Y-Y (2007) The high-dimension, low-sample-size geometric representation holds under mild conditions. Biometrika 94:760–766
2. Aoshima M (2005) Statistical inference in two-stage sampling. Amer Math Soc Trans 215:125–145
3. Aoshima M, Mukhopadhyay N (1998) Fixed-width simultaneous confidence intervals for multinormal means in several intraclass correlation models. J Mult Anal 66(1):46–63
4. Aoshima M, Takada Y (2004) Asymptotic second-order efficiency for multivariate two-stage estimation of a linear function of normal mean vectors. Seq Anal 23(3):333–353

5. Aoshima M, Takada Y, Srivastava MS (2002) A two-stage procedure for estimating a linear function of k multinormal mean vectors when covariance matrices and unknown. J Stat Plan Infer 100:109–119
6. Aoshima M, Yata K (2010) Asymptotic second-order consistency for two-stage estimation methodologies and its applications. Ann Inst Statist Math 62:571–600
7. Aoshima M, Yata K (2011) Two-stage procedures for high-dimensional data. Seq Anal (Editor's special invited paper) 30:356-399
8. Bai Z, Sarandasa H (1996) Effect of high dimension: By an example of a two sample problem. Statistica Sinica 6:311–329
9. Bradley RC (2005) Basic properties of strong mixing conditions. A survey and some open questions. Probab Surv 2:107–144(electronic)
10. 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
11. Ghosh M, Mukhopadhyay N, Sen PK (1997) Sequential estimation. Wiley, New York
12. Hall P, Marron JS, Neeman A (2005) Geometric representation of high dimension, low sample size data. J R Statist Soc Ser B 67:427–444
13. Kolmogorov AN, Rozanov YA (1960) On strong mixing conditions for stationary gaussian processes. Theory Probab Appl 5:204–208
14. Mukhopadhyay N, Duggan WT (1997) Can a two-stage procedure enjoy second-order properties? Sankhyā Ser A 59:435–448
15. Mukhopadhyay N, Duggan WT (1999) On a two-stage procedure having second-order properties with applications. Ann Inst Statist Math 51:621–636
16. Pollard KS, Dudoit S, van der Laan, MJ (2005) Multiple testing procedures: R multtest package and applications to genomics. In: Gentleman R, Carey V, Huber W, Irizarry R, Dudoit S (ed) Bioinformatics and computational biology solutions using R and bioconductor. Springer, New York, pp 249–271
17. Srivastava MS (2005) Some tests concerning the covariance matrix in high dimensional data J Japan Statist Soc 35:251–272
18. Stein C (1945) A two-sample test for a linear hypothesis whose power is independent of the variance, Ann Math Statist 16:243–258
19. Yata K (2010) Effective two-stage estimation for a linear function of high-dimensional gaussian means. Seq Anal 29:463–482
20. Yata K, Aoshima M (2009a) Double shrink methodologies to determine the sample size via covariance structures. J Stat Plan Infer 139:81–99
21. Yata K, Aoshima M (2009b) PCA consistency for non-gaussian data in high dimension, low sample size context. Commun Statist-Theory Meth, Special Issue Honoring Zacks S (ed Mukhopadhyay N) 38:2634–2652
22. Yata K, Aoshima M (2010a) Effective PCA for high-dimension, low-sample-size data with singular value decomposition of cross data matrix. J Mult Anal 101:2060–2077
23. Yata K, Aoshima M (2010b) Intrinsic dimensionality estimation of high dimension, low sample size data with d-asymptotics. Commun Statist-Theory Meth, Special Issue Honoring Akahira M (ed Aoshima M) 39:1511–1521
24. Yata K, Aoshima M (2012) Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations. J Mult Anal 105:193–215