High-dimensional multiple comparison procedures among mean vectors under covariance heterogeneity

Masashi Hyodo[a], Takahiro Nishiyama[b], Hiromasa Hayashi[c]

[a] *Faculty of Economics, Kanagawa University,*
3-27-1 Rokkakubashi, Kanagawa-ku, Yokohama-shi, Kanagawa, Japan.
E-Mail: `caicmhy@gmail.com`
[b] *Department of Business Administration, Senshu University,*
2-1-1, Higashimita, Tama-ku, Kawasaki-shi, Kanagawa 214-8580, Japan.
E-Mail:`nishiyama@isc.senshu-u.ac.jp`
[c] *Department of Mathematical Sciences, Graduate School of Engineering, Osaka Prefecture University,*
1-1 Gakuen-cho, Naka-ku, Sakai, Osaka 599-8531, Japan.
E-Mail:`bsk31h@gmail.com`

Abstract

In this paper, we discuss two typical multivariate multiple comparisons procedures among mean vectors: that is, pairwise comparisons and comparisons with a control. In traditional multivariate analysis, these multivariate multiple comparisons procedures are constructed based on Hotelling's T^2 statistic in multivariate normal populations. However, in high-dimensional settings, such when the dimensions exceed total sample sizes, these methods cannot be applied. In such cases, Takahashi et al. (2013) proposed asymptotically conservative simultaneous confidence intervals under the assumption of homogeneity of variance-covariance matrices across groups. Unfortunately, these simultaneous confidence intervals are not asymptotically conservative when this assumption is violated. Motivated by this point, we newly obtain asymptotically conservative confidence intervals based on L^2-type statistic without assuming that the variance-covariance matrices are homogeneous across groups. Empirical results indicate that the proposed simultaneous confidence intervals outperform existing procedures.

AMS 2000 subject classification: Primary 62H15; secondary 62F03.

Key words: Comparisons with a control, Covariance heterogeneity, High-dimensional data, Multiple comparisons, Pairwise comparisons.

1. Introduction

The study of multiple comparisons under univariate and multivariate analyses has been undertaken by many authors, see, e.g., Hochberg and Tamhane (1987), Hsu (1996) and Bretz et al. (2010). In this paper, we discuss two typical multivariate multiple comparisons procedures among mean vectors: that is, pairwise comparisons and comparisons with a control. When we consider multivariate multiple comparisons among mean vectors, we usually deal with simultaneous confidence intervals. So, it is well established that constructing simultaneous confidence intervals among mean vectors is important for this problem.

Let $\mathbf{x}_{ij}$ for $i \in \{1, \ldots, k\}$ and $j \in \{1, \ldots, n_i\}$ be independently distributed as the p-dimensional normal distribution with mean vector $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$, which is denoted as $\mathcal{N}_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$. Besides, let $\mathbb{R}^p_* = \mathbb{R}^p \setminus \{\mathbf{0}\}$. Then, we consider simultaneous confidence intervals for pairwise multiple comparisons among mean vectors, that is, for the set of all linear combinations of the mean difference $\mathbf{a}^\top(\boldsymbol{\mu}_\ell - \boldsymbol{\mu}_m) = \mathbf{a}^\top\boldsymbol{\delta}_{\ell m}$ for all $\mathbf{a} \in \mathbb{R}^p_*$ and for all $\ell, m \in \{1, \ldots, k\}$. Also, letting the first population be a control, we consider simultaneous confidence intervals for comparisons with a control, that is, for the set of all linear combinations of the mean difference $\mathbf{a}^\top(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_m) = \mathbf{a}^\top\boldsymbol{\delta}_{1m}$ for all $\mathbf{a} \in \mathbb{R}^p_*$ and for all $m \in \{2, \ldots, k\}$.

In general, it is difficult to construct so-called exact simultaneous confidence intervals in which the nominal confidence level and coverage probability match. Thus, the conservative simultaneous confidence intervals in which coverage probability is larger than nominal confidence level is often studied. When $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_k$ and $p \leq n - k$ where $n = \sum_{i=1}^{k} n_i$, it is well known that simultaneous confidence intervals for pairwise multiple comparisons and comparisons with a control among mean vectors are based on Hotelling's T^2 statistic. That has been extensively studied by many statisticians, see, e.g., Seo and Siotani (1992), Seo, Mano and Fujikoshi (1994), and Seo and Nishiyama (2008).

Recently, high-dimensional data are frequently collected in various research and industrial areas. For high-dimensional settings such as $p > n - k$, the sample covariance matrix becomes singular, and hence, Hotelling's T^2 statistic cannot be defined. In these situations, by changing T^2 statistic to Dempster's (1958, 1960) statistics, Hyodo et al. (2014) proposed simultaneous confidence intervals for multiple comparisons among mean vectors in high-dimensional settings with a balanced sample case. Also, Takahashi et al. (2013) offered an extension of the results with a balanced sample case by Hyodo et al. (2014) to an unbalanced sample case.

Also, in recent year, testing procedures for high-dimensional data which tests the equality of mean vectors under covariance heterogeneity have been paid much attention. For example, Chen and Qin (2010) proposed an L^2-type statistic for two sample test without assuming the equality of two covariance matrices, that is, multivariate Behrens-Fisher problem. Besides, other important testing procedures under covariance heterogeneity have been studied by many authors, see, e.g., Aoshima and Yata (2011), Nishiyama et al. (2013), Feng et al. (2015), Hu et al. (2017), Ishii et al. (2019) and Zhang et al. (2021).

In this paper, we discuss multivariate multiple comparisons procedures among mean vectors. For this problem, as mentioned above, Hyodo et al. (2014) and Takahashi et al. (2013) assumed $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_k$ to construct simultaneous confidence intervals. Unfortunately, when $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_k$ is violated, these simultaneous confidence intervals are not asymptotically conservative (for details, we state in section 2). Motivated by this point, we newly propose a pairwise multiple comparisons and comparisons with a control among mean vectors based on the following L^2-type statistic without assuming that $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_k$:

$$\widetilde{H}_{\ell m} = \|\widehat{\boldsymbol{\delta}}_{\ell m} - \boldsymbol{\delta}_{\ell m}\|^2 - \frac{\mathrm{tr}(\mathbf{S}_\ell)}{n_\ell} - \frac{\mathrm{tr}(\mathbf{S}_m)}{n_m},$$

where $\widehat{\boldsymbol{\delta}}_{\ell m} = \overline{\mathbf{x}}_\ell - \overline{\mathbf{x}}_m$ for $\ell, m \in \{1, \ldots, k\}$, and $\overline{\mathbf{x}}_i = n_i^{-1} \sum_{j=1}^{n_i} \mathbf{x}_{ij}$ is the i-th sample mean vector and $\mathbf{S}_i = (n_i - 1)^{-1} \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \overline{\mathbf{x}}_i)(\mathbf{x}_{ij} - \overline{\mathbf{x}}_i)^\top$ is the i-th sample covariance matrix for $i \in \{1, \ldots, k\}$. Chen and Qin (2010) showed asymptotic normality of this statistic. This fact also provides asymptotic validity to using percentage points of standard normal distribution $\mathcal{N}(0, 1)$ as an approximation for percentage points of the L^2-type statistic in high-dimensional settings. In this paper, we simply call this approximation a 'normal approximation'. However, the normal approximation is often too loose or fails to capture the tail behavior of the resulting distribution. For this reason, we newly derive an Edgeworth expansion and Cornish-Fisher expansion for studentized L^2-type statistic and construct a confidence interval by applying the Cornish-Fisher expansion. We also show that asymptotic coverage probability is greater than or equal to the nominal confidence level (that is, asymptotically conservative).

The remainder of this paper is organized as follows: In section 2, we investigate the effect of heteroscedasticity after introducing the simultaneous confidence intervals of Takahashi et al. (2013). In section 3, we derive an Edgeworth expansion and Cornish-Fisher expansion of studentized L^2-type statistic. Also, based on these results, we construct new simultaneous confidence intervals for pairwise multiple comparisons and comparisons with a control among mean vectors without assuming that $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_k$. In section 4, via Monte Carlo simulations, we compare our proposed simultaneous confidence intervals with existing simultaneous confidence intervals given by Takahashi et al. (2013) and conclude with advantages of the proposed procedures. Further, to illustrate our results, we present a real data analysis. Finally, we provide some concluding remarks. Proofs of theorems and lemmas are detailed in the appendix.

2. Introduction to previous studies and the effect of covariance heterogeneity

2.1. Introduction to previous studies

Let the pooled sample covariance matrix be

$$\mathbf{S} = \frac{1}{n-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\mathbf{x}_{ij} - \overline{\mathbf{x}}_i)(\mathbf{x}_{ij} - \overline{\mathbf{x}}_i)^\top,$$

where $n = \sum_{i=1}^{k} n_i$. Dempster (1958, 1960) proposed the following statistic:

$$\widetilde{D}_{\ell m} = w_{\ell m}^{-1}\frac{\|\widehat{\boldsymbol{\delta}}_{\ell m} - \boldsymbol{\delta}_{\ell m}\|^2}{\mathrm{tr}(\mathbf{S})},$$

where $w_{\ell m} = 1/n_\ell + 1/n_m$ for $\ell, m \in \{1, \ldots, k\}$. We note that $\widetilde{D}_{\ell m}$ can be clearly defined even if $p > n - k$. When $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}_0$, the asymptotic mean and asymptotic variance of $\widetilde{D}_{\ell m}$ are given by

$$\mathrm{E}(\widetilde{D}_{\ell m}) \approx 1,\ \mathrm{var}(\widetilde{D}_{\ell m}) \approx \frac{2\mathrm{tr}(\boldsymbol{\Sigma}_0^2)}{\{\mathrm{tr}(\boldsymbol{\Sigma}_0)\}^2} =: \sigma^2.$$

To construct simultaneous confidence intervals, Takahashi et al. (2013) defined so-called studentized statistic

$$D_{\ell m} = \frac{1}{\widehat{\sigma}}\left\{w_{\ell m}^{-1}\frac{\|\widehat{\boldsymbol{\delta}}_{\ell m} - \boldsymbol{\delta}_{\ell m}\|^2}{\mathrm{tr}(\mathbf{S})} - 1\right\},$$

for $\ell \neq m,\ \ell, m \in \{1, \ldots, k\}$ and

$$\widehat{\sigma} = \frac{1}{\mathrm{tr}(\mathbf{S})}\sqrt{\frac{2(n-k)^2}{(n-k+2)(n-k-1)}\left\{\mathrm{tr}(\mathbf{S}^2) - \frac{\{\mathrm{tr}(\mathbf{S})\}^2}{(n-k)}\right\}}.$$

Let nominal confidence level be $1-\alpha$, $\alpha \in (0, 1)$. Next, Takahashi et al. (2013) considered simultaneous confidence intervals for pairwise multiple comparisons and comparisons with a control, respectively, consisting of the following:

$$\left[\mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{\ell m} - \mathrm{D}_{\mathrm{pw}}^{\ell m}, \mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{\ell m} + \mathrm{D}_{\mathrm{pw}}^{\ell m}\right],\ \forall \mathbf{a} \in \mathbb{R}_*^p,\ \forall \ell < m,\ \ell, m \in \{1, \ldots, k\},$$
$$\left[\mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{1m} - \mathrm{D}_{\mathrm{c}}^{1m}, \mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{1m} + \mathrm{D}_{\mathrm{c}}^{1m}\right],\ \forall \mathbf{a} \in \mathbb{R}_*^p,\ \forall m \in \{2, \ldots, k\},$$

where

$$\mathrm{D}_{\mathrm{pw}}^{\ell m} = \|\mathbf{a}\|\sqrt{w_{\ell m}\mathrm{tr}(\mathbf{S})(1 + \widehat{\sigma}d_{\mathrm{pw}})}\ ,\ \mathrm{D}_{\mathrm{c}}^{1m} = \|\mathbf{a}\|\sqrt{w_{1m}\mathrm{tr}(\mathbf{S})(1 + \widehat{\sigma}d_{\mathrm{c}})}\ .$$

Here, exact critical values d_{pw} and d_{c} satisfy as follows:

$$\Pr\left(\max_{1\le\ell<m\le k} D_{\ell m} \le d_{\mathrm{pw}}\right) = 1 - \alpha,\ \Pr\left(\max_{2\le m\le k} D_{1m} \le d_{\mathrm{c}}\right) = 1 - \alpha.$$

Because it is difficult to obtain exact critical values for d_{pw} and d_{c} in simultaneous confidence intervals, Bonferroni's approximate procedure is discussed by Takahashi et al. (2013). By using Bonferroni's inequality, coverage probabilities of the two confidence intervals based on Dempster's statistic can be evaluated as

$$\Pr\left(\max_{1\le\ell<m\le k} D_{\ell m} \le d_{\mathrm{pw}}\right) \ge 1 - \sum_{1\le\ell<m\le k}\Pr\left(D_{\ell m} \ge d_{\mathrm{pw}}\right),$$
$$\Pr\left(\max_{2\le m\le k} D_{1m} \le d_{\mathrm{c}}\right) \ge 1 - \sum_{2\le m\le k}\Pr\left(D_{1m} \ge d_{\mathrm{c}}\right),$$

respectively. Further, Takahashi et al. (2013) constructed an asymptotically conservative simultaneous confidence interval by choosing d_{pw} and d_{c} so that $\Pr(D_{\ell m} \ge d_{\mathrm{pw}}) = \alpha/K_{\mathrm{pw}} + o(1)$ and $\Pr(D_{1m} \ge d_{\mathrm{c}}) = \alpha/K_{\mathrm{c}} + o(1)$, where $K_{\mathrm{pw}} = k(k-1)/2$ and $K_{\mathrm{c}} = k - 1$. The specific forms of these confidence intervals are obtained by following.

1. Simultaneous confidence intervals for pairwise multiple comparisons among mean vectors are given by

$$TCI_{\mathrm{pw1}} = \left[\mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{\ell m} - \mathrm{D}_{1\mathrm{pw}}^{\ell m}, \mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{\ell m} + \mathrm{D}_{1\mathrm{pw}}^{\ell m}\right],\ \forall \mathbf{a} \in \mathbb{R}_*^p,\ \forall \ell < m,\ \ell, m \in \{1, \dots, k\}, \tag{2.1}$$

where

$$\mathrm{D}_{1\mathrm{pw}}^{\ell m} = \|\mathbf{a}\| \sqrt{w_{\ell m}\mathrm{tr}(\mathbf{S})(1 + \widehat{\sigma} z_{\alpha_{\mathrm{pw}}})}.$$

Here, $\alpha_{\mathrm{pw}} = \alpha/K_{\mathrm{pw}}$ and z_a denotes the upper $100 \times a$ percentile of the standard normal distribution $\mathcal{N}(0, 1)$.

2. Simultaneous confidence intervals for multiple comparisons with a control among mean vectors are given by

$$TCI_{\mathrm{c1}} = \left[\mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{1m} - \mathrm{D}_{1\mathrm{c}}^{1m}, \mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{1m} + \mathrm{D}_{1\mathrm{c}}^{1m}\right],\ \forall \mathbf{a} \in \mathbb{R}_*^p,\ \forall m \in \{2, \dots, k\}, \tag{2.2}$$

where

$$\mathrm{D}_{1\mathrm{c}}^{1m} = \|\mathbf{a}\| \sqrt{w_{1m}\mathrm{tr}(\mathbf{S})(1 + \widehat{\sigma} z_{\alpha_{\mathrm{c}}})}.$$

3. Simultaneous confidence intervals for pairwise multiple comparisons among mean vectors are given by

$$TCI_{\mathrm{pw2}} = \left[\mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{\ell m} - \mathrm{D}_{2\mathrm{pw}}^{\ell m}, \mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{\ell m} + \mathrm{D}_{2\mathrm{pw}}^{\ell m}\right],\ \forall \mathbf{a} \in \mathbb{R}_*^p,\ \forall \ell < m,\ \ell, m \in \{1, \dots, k\}, \tag{2.3}$$

where

$$\mathrm{D}_{2\mathrm{pw}}^{\ell m} = \|\mathbf{a}\| \sqrt{w_{\ell m}\mathrm{tr}(\mathbf{S})\left\{1 + \widehat{\sigma}\widehat{d}(z_{\alpha_{\mathrm{pw}}})\right\}}.$$

Here, $\widehat{d}(x)$ is estimated using the Cornish-Fisher expansion, which is defined by

$$\widehat{d}(x) = x + \frac{1}{\sqrt{p}}\left(\frac{\sqrt{2}\widehat{c_3}}{3\sqrt{\widehat{c_2}^3}}\right)(x^2 - 1) + \frac{1}{p}\left\{\frac{\widehat{c_4}}{2\widehat{c_2}^2}x(x^2 - 3) - \frac{2\widehat{c_3}^2}{9\widehat{c_2}^3}x(2x^2 - 5)\right\} + \frac{1}{2n}x,$$

where

$$\begin{aligned}
\widehat{c_1} &= \frac{\mathrm{tr}(\mathbf{S})}{p},\ \widehat{c_2} = \frac{n^2}{(n+2)(n-1)p}\left[\mathrm{tr}(\mathbf{S}^2) - \frac{\{\mathrm{tr}(\mathbf{S})\}^2}{n}\right],\\
\widehat{c_3} &= \frac{n^4}{(n+4)(n+2)(n-1)(n-2)p}\left[\mathrm{tr}(\mathbf{S}^3) - \frac{3\mathrm{tr}(\mathbf{S}^2)\mathrm{tr}(\mathbf{S})}{n} + \frac{2\{\mathrm{tr}(\mathbf{S})\}^3}{n^2}\right],\\
\widehat{c_4} &= \frac{n^3}{(n+6)(n+4)(n+2)(n+1)(n-1)(n-2)(n-3)p}\\
&\quad \times \Big[n^2(n^2+n+2)\mathrm{tr}(\mathbf{S}^4) - 4n(n^2+n+2)\mathrm{tr}(\mathbf{S}^3)\mathrm{tr}(\mathbf{S})\\
&\quad - n(2n^2+3n-6)\{\mathrm{tr}(\mathbf{S}^2)\}^2 + 2n(5n+6)\mathrm{tr}(\mathbf{S}^2)\{\mathrm{tr}(\mathbf{S})\}^2 - (5n+6)\{\mathrm{tr}(\mathbf{S})\}^4\Big].
\end{aligned}$$

4. Simultaneous confidence intervals for multiple comparisons with a control among mean vectors are given by

$$TCI_{\mathrm{c2}} = \left[\mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{1m} - \mathrm{D}_{2\mathrm{c}}^{1m}, \mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{1m} + \mathrm{D}_{2\mathrm{c}}^{1m}\right],\ \forall \mathbf{a} \in \mathbb{R}_*^p,\ \forall m \in \{2, \dots, k\}, \tag{2.4}$$

where

$$\mathrm{D}_{2\mathrm{c}}^{1m} = \|\mathbf{a}\| \sqrt{w_{1m}\mathrm{tr}(\mathbf{S})\left\{1 + \widehat{\sigma}\widehat{d}(z_{\alpha_{\mathrm{c}}})\right\}}.$$

Simultaneous confidence intervals given by 1 and 2 are constructed using percentage points of the limit distribution of $D_{\ell m}$. Simultaneous confidence intervals given by 3 and 4 are constructed using an estimated Cornish-Fisher expansion for Dempster statistic $D_{\ell m}$.

2.2. The effect of covariance heterogeneity

In this section, we discuss the effect of covariance heterogeneity on simultaneous confidence intervals based on Dempster's statistic $\widetilde{D}_{\ell m}$ when the assumption $\mathbf{\Sigma}_1 = \cdots = \mathbf{\Sigma}_k$ is violated. We assume the following two conditions for asymptotic assessment.

(A1) $p \to \infty,\ n_0 = \min\{n_1, \ldots, n_k\} \to \infty,\ lim_{n_0, p \to \infty} p/n_0 \in (0, \infty)$, and $lim_{n_0 \to \infty} n_i/n_0 \in (0, \infty)$ for $i \in \{1, \ldots, k\}$.

(A2) For any $i \in \{1, \ldots, k\}$, the eigenvalues of $\mathbf{\Sigma}_i$ admit the representation

$$\lambda_r(\mathbf{\Sigma}_i) = a_{i(r)} p^{\beta_{i(r)}},\ r \in \{1, \ldots, t_i\},\ \lambda_r(\mathbf{\Sigma}_i) = c_{i(r)},\ r \in \{t_i + 1, \ldots, p\},$$

where $a_{i(r)},\ c_{i(r)}$ and $\beta_{i(r)}$ are positive and fixed constants and t_i is a fixed positive integer. Further, $\beta_{(1)} = \max\{\beta_{1(1)}, \ldots, \beta_{k(1)}\} < 1/2$.

From Takahashi et al. (2013), under (A1), (A2), and $\mathbf{\Sigma}_1 = \cdots = \mathbf{\Sigma}_k$, all simultaneous confidence intervals are asymptotically conservative. When $\mathbf{\Sigma}_1 = \cdots = \mathbf{\Sigma}_k$ is violated, we will show that asymptotic conservatism does not hold using a simple example. Beforehand, we will prepare the following supplementary lemma.

Lemma 1. *Under (A1) and (A2), $\widetilde{D}_{\ell m} = m^*_{\ell m} + o_p(1)$ and $\widehat{\sigma} = o_p(1)$, where*

$$m^*_{\ell m} = \left\{ \frac{n_m}{n_{\ell m}} \mathrm{tr}(\mathbf{\Sigma}_\ell) + \frac{n_\ell}{n_{\ell m}} \mathrm{tr}(\mathbf{\Sigma}_m) \right\} / \sum_{i=1}^{k} (n_i - 1)/(n - k) \mathrm{tr}(\mathbf{\Sigma}_i).$$

Here, $n_{\ell m} = n_\ell + n_m$.

Proof. See, Appendix A. □

As a simple example of violating the assumption $\mathbf{\Sigma}_1 = \cdots = \mathbf{\Sigma}_k$, we consider $\mathbf{\Sigma}_i = (k-i+1)\mathbf{\Sigma}_0$ for all $i \in \{1, \ldots, k\}$. We also assume $n_i = n_0$. Then $m^*_{\ell m} = \{2(k+1) - \ell - m\}/(k+1)$ and $1 - m^*_{12} = -(k-2)/(k+1)$. By using Lemma 1, under (A1) and (A2), for any $z \in \mathbb{R}$ and any number $k > 2$,

$$\begin{aligned} \Pr(D_{12} \le z) =& \Pr\{\widetilde{D}_{12} - (2k-1)/(k+1) \le -(k-2)/(k+1) + \widehat{\sigma} z\} \\ \le& \Pr\{|\widetilde{D}_{12} - (2k-1)/(k+1)| > (k-2)/(k+1) - \widehat{\sigma} z\} \\ =& \Pr\{|\widetilde{D}_{12} - (2k-1)/(k+1)| > (k-2)/(k+1)\} + o(1) = o(1). \end{aligned} \quad (2.5)$$

Also, coverage probability for each simultaneous confidence intervals TCI_{pw1} and TCI_{c1} are evaluated as

$$\Pr\left(\max_{1 \le \ell < m \le k} D_{\ell m} \le z_{\alpha_{\mathrm{pw}}} \right) \le \Pr\left(D_{12} \le z_{\alpha_{\mathrm{pw}}} \right),\ \Pr\left(\max_{2 \le m \le k} D_{1m} \le z_{\alpha_{\mathrm{c}}} \right) \le \Pr\left(D_{12} \le z_{\alpha_{\mathrm{c}}} \right). \quad (2.6)$$

From (2.5) and (2.6), coverage probability for each confidence interval convergence to 0, that is, asymptotically conservative, does not hold. From this simple example, we consider that Takahashi et al. (2013)'s simultaneous confidence intervals do not always become asymptotically conservative when $\mathbf{\Sigma}_1 = \cdots = \mathbf{\Sigma}_k$ is violated. Since this phenomenon is essentially caused by the deviation of asymptotic mean of a Dempster statistic from 1, other statistics should be considered for the construction of confidence intervals when $\mathbf{\Sigma}_1 = \cdots = \mathbf{\Sigma}_k$ is violated.

3. Main results

3.1. Asymptotic results for studentized L^2-type statistic

As explained in the previous section, Dempster's statistic is not suitable when the covariance has heterogeneity. To deal with a case of covariance heterogeneity, we utilize the L^2-type statistic defined below.

$$\widetilde{H}_{\ell m} = \|\widehat{\boldsymbol{\delta}}_{\ell m} - \boldsymbol{\delta}_{\ell m}\|^2 - \frac{\mathrm{tr}(\mathbf{S}_\ell)}{n_\ell} - \frac{\mathrm{tr}(\mathbf{S}_m)}{n_m}.$$

The mean and variance of this statistic are as follows:

$$\mathrm{E}(\widetilde{H}_{\ell m}) = 0,\ \mathrm{var}(\widetilde{H}_{\ell m}) = \sum_{g\in\{\ell,m\}} \frac{2\mathrm{tr}(\boldsymbol{\Sigma}_g^2)}{n_g(n_g-1)} + \frac{4\mathrm{tr}(\boldsymbol{\Sigma}_\ell\boldsymbol{\Sigma}_m)}{n_\ell n_m} =: \sigma_{\ell m}^2.$$

From this result, we note that $\widetilde{H}_{\ell m}$ is suitable because the expectation is 0 even when the covariance has heterogeneity, meaning it is unbiased. We define a so-called studentized statistic with the standard deviation $\sigma_{\ell m}$ replaced by an estimator for application to simultaneous confidence intervals:

$$H_{\ell m} = \frac{\|\widehat{\boldsymbol{\delta}}_{\ell m} - \boldsymbol{\delta}_{\ell m}\|^2 - \mathrm{tr}(\mathbf{S}_\ell)/n_\ell - \mathrm{tr}(\mathbf{S}_m)/n_m}{\widehat{\sigma}_{\ell m}},$$

where

$$\widehat{\sigma}_{\ell m} = \sqrt{\sum_{g\in\{\ell,m\}} \frac{2(n_g-1)}{n_g(n_g+1)(n_g-2)}\left[\mathrm{tr}(\mathbf{S}_g^2) - \frac{\{\mathrm{tr}(\mathbf{S}_g)\}^2}{n_g-1}\right] + \frac{4\mathrm{tr}(\mathbf{S}_\ell\mathbf{S}_m)}{n_\ell n_m}}.$$

First, we derive the Edgeworth expansion of studentized L^2-type statistics in the following lemma.

Lemma 2. *Under (A1) and (A2), for any x in the compact subset of $\mathbb{R}$,*

$$\Pr(H_{\ell m} \le x) = \Phi(x) + \frac{4b_{\ell m}}{3\sigma_{\ell m}^3}(1-x^2)\phi(x) + o(p^{\beta_{(1)}-1/2}), \tag{3.1}$$

where

$$b_{\ell m} = \sum_{g\in\{\ell,m\}} \frac{(n_g-2)\mathrm{tr}(\boldsymbol{\Sigma}_g^3)}{n_g^2(n_g-1)^2} + \frac{3\mathrm{tr}(\boldsymbol{\Sigma}_\ell^2\boldsymbol{\Sigma}_m)}{n_\ell^2 n_m} + \frac{3\mathrm{tr}(\boldsymbol{\Sigma}_\ell\boldsymbol{\Sigma}_m^2)}{n_\ell n_m^2}.$$

Proof. See, Appendix B. □

Using Lemma 2, under (A1) and (A2), for any x in the compact subset of $\mathbb{R}$,

$$\Pr(H_{\ell m} \le x) = \Phi(x) + O(p^{\beta_{(1)}-1/2}) = \Phi(x) + o(1). \tag{3.2}$$

Thus, we can see the asymptotic normality of $H_{\ell m}$ and its convergence rate $O(p^{\beta_{(1)}-1/2})$. This also provides asymptotic validity for using the percentage points of $\mathcal{N}(0,1)$ as an approximation for those of $H_{\ell m}$ in high-dimensional settings.

Next, we consider an approximate percentage point that improves convergence rate $O(p^{\beta_{(1)}-1/2})$. Specifically, we derive the so-called Cornish-Fisher expansion, which is a correction of normal approximation. We obtain the Cornish-Fisher expansion:

$$cf_{\ell m}(x) = x + \frac{4b_{\ell m}}{3\sigma_{\ell m}^3}(x^2-1).$$

By using the result (2.2) in Hall (1983) along with Lemma 2, under (A1) and (A2), for any x in the compact subset of $\mathbb{R}$,

$$\Pr\{H_{\ell m} \le cf_{\ell m}(x)\} = \Phi(x) + o(p^{\beta_{(1)}-1/2}).$$

Thus, we confirm that the convergence rate of $cf_{\ell m}(x)$ improves the convergence rate of normal approximation. However, since $cf_{\ell m}(x)$ contains unknown parameters $\sigma_{\ell m}$ and $b_{\ell m}$, we need to estimate $cf_{\ell m}(x)$.

So, finally, we consider estimation of the Cornish-Fisher expansion $cf_{\ell m}(x)$. The unbiased estimator of $b_{\ell m}$ is given

$$\begin{aligned}\widehat{b}_{\ell m} = &\sum_{g\in\{\ell,m\}} \frac{(n_g-1)^2}{(n_g-3)n_g^2(n_g+1)(n_g+3)}\left[\mathrm{tr}(\mathbf{S}_g^3) - \frac{3\mathrm{tr}(\mathbf{S}_g^2)\mathrm{tr}(\mathbf{S}_g)}{n_g-1} + \frac{2\{\mathrm{tr}(\mathbf{S}_g)\}^3}{(n_g-1)^2}\right]\\ &+ \frac{3(n_\ell-1)^2}{(n_\ell-2)n_\ell^2(n_\ell+1)n_m}\left\{\mathrm{tr}(\mathbf{S}_\ell^2\mathbf{S}_m) - \frac{\mathrm{tr}(\mathbf{S}_\ell\mathbf{S}_m)\mathrm{tr}(\mathbf{S}_\ell)}{n_\ell-1}\right\}\\ &+ \frac{3(n_m-1)^2}{(n_m-2)n_m^2(n_m+1)n_\ell}\left\{\mathrm{tr}(\mathbf{S}_m^2\mathbf{S}_\ell) - \frac{\mathrm{tr}(\mathbf{S}_m\mathbf{S}_\ell)\mathrm{tr}(\mathbf{S}_m)}{n_m-1}\right\}.\end{aligned}$$

Properties of estimators $\widehat{\sigma}^2_{\ell m}$ and $\widehat{b}_{\ell m}$ are summarized in the following lemma.

Lemma 3. *$\mathrm{E}(\widehat{\sigma}^2_{\ell m}) = \sigma^2_{\ell m}$ and $\mathrm{E}(\widehat{b}_{\ell m}) = b_{\ell m}$. Also, under (A1) and (A2), $\widehat{\sigma}^2_{\ell m}/\sigma_{\ell m} = 1 + o_p(1)$ and $\widehat{b}_{\ell m}/b_{\ell m} = 1 + o_p(1)$.*

Proof. See, Appendix C. □

By replacing $\sigma_{\ell m}$ and $b_{\ell m}$ contained in $cf_{\ell m}(x)$ with their estimators $\widehat{\sigma}_{\ell m}$ and $\widehat{b}_{\ell m}$, we obtain $\widehat{cf}_{\ell m}(x)$. Also, the asymptotic property of the estimated Cornish-Fisher expansion $\widehat{cf}_{\ell m}(x)$ is given in the following theorem.

Theorem 1. *Under (A1) and (A2), for any point x in the compact subset of $\mathbb{R}$,*

$$\Pr\{H_{\ell m} \le \widehat{cf}_{\ell m}(x)\} = \Phi(x) + o(p^{\beta_{(1)}-1/2}).$$

Proof. See, Appendix D. □

3.2. Simultaneous confidence intervals

In this section, we construct simultaneous confidence intervals based on statistic $H_{\ell m}$ that is valid without assuming $\mathbf{\Sigma}_1 = \cdots = \mathbf{\Sigma}_k$. We define the nominal confidence level as $1-\alpha$, $\alpha \in (0,1)$. Let $h^{\ell m}_{\mathrm{pw}}$ and h^{1m}_{c} be exact critical values satisfy

$$\Pr\left[\bigcap_{1\le \ell<m\le k}\left\{H_{\ell m} \le h^{\ell m}_{\mathrm{pw}}\right\}\right] = 1-\alpha,\ \ \Pr\left[\bigcap_{2\le m\le k}\left\{H_{1m} \le h^{1m}_{\mathrm{c}}\right\}\right] = 1-\alpha.$$

And let

$$P_{\mathrm{pw}} = \Pr\left[\bigcap_{1\le \ell<m\le k}\ \bigcap_{\mathbf{a}\in\mathbb{R}^p_*}\left\{|\mathbf{a}^\top(\widehat{\boldsymbol{\delta}}_{\ell m} - \boldsymbol{\delta}_{\ell m})| \le \|\mathbf{a}\|\sqrt{\frac{\mathrm{tr}(\mathbf{S}_\ell)}{n_\ell} + \frac{\mathrm{tr}(\mathbf{S}_m)}{n_m} + \widehat{\sigma}_{\ell m}h^{\ell m}_{\mathrm{pw}}}\right\}\right],$$
$$P_{\mathrm{c}} = \Pr\left[\bigcap_{2\le m\le k}\ \bigcap_{\mathbf{a}\in\mathbb{R}^p_*}\left\{|\mathbf{a}^\top(\widehat{\boldsymbol{\delta}}_{1m} - \boldsymbol{\delta}_{1m})| \le \|\mathbf{a}\|\sqrt{\frac{\mathrm{tr}(\mathbf{S}_1)}{n_1} + \frac{\mathrm{tr}(\mathbf{S}_m)}{n_m} + \widehat{\sigma}_{1m}h^{1m}_{\mathrm{c}}}\right\}\right].$$

Then we can evaluate P_{pw} as follows.

$$\begin{aligned} P_{\mathrm{pw}} &= \Pr\left[\bigcap_{1\le \ell<m\le k}\left\{\max_{\mathbf{a}\in\mathbb{R}^p_*}\frac{|\mathbf{a}^\top(\widehat{\boldsymbol{\delta}}_{\ell m} - \boldsymbol{\delta}_{\ell m})|^2}{\|\mathbf{a}\|^2} \le \frac{\mathrm{tr}(\mathbf{S}_\ell)}{n_\ell} + \frac{\mathrm{tr}(\mathbf{S}_m)}{n_m} + \widehat{\sigma}_{\ell m}h^{\ell m}_{\mathrm{pw}}\right\}\right] \\ &= \Pr\left[\bigcap_{1\le \ell<m\le k}\left\{\|\widehat{\boldsymbol{\delta}}_{\ell m} - \boldsymbol{\delta}_{\ell m}\|^2 \le \frac{\mathrm{tr}(\mathbf{S}_\ell)}{n_\ell} + \frac{\mathrm{tr}(\mathbf{S}_m)}{n_m} + \widehat{\sigma}_{\ell m}h^{\ell m}_{\mathrm{pw}}\right\}\right] \\ &= \Pr\left[\bigcap_{1\le \ell<m\le k}\left\{H_{\ell m} \le h^{\ell m}_{\mathrm{pw}}\right\}\right] = 1-\alpha. \end{aligned}$$

Also, using same strategy, we can evaluate P_{c} as follows.

$$P_{\mathrm{c}} = \Pr\left[\bigcap_{2\le m\le k}\left\{\max_{\mathbf{a}\in\mathbb{R}^p_*}\frac{|\mathbf{a}^\top(\widehat{\boldsymbol{\delta}}_{1m} - \boldsymbol{\delta}_{1m})|^2}{\|\mathbf{a}\|^2} \le \frac{\mathrm{tr}(\mathbf{S}_1)}{n_1} + \frac{\mathrm{tr}(\mathbf{S}_m)}{n_m} + \widehat{\sigma}_{1m}h^{1m}_{\mathrm{c}}\right\}\right] = 1-\alpha.$$

Therefore, we can obtain simultaneous confidence intervals for pairwise multiple comparisons and comparisons with a control, respectively, consisting of the following:

$$\left[\mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{\ell m} - \mathrm{H}^{\ell m}_{\mathrm{pw}},\ \mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{\ell m} + \mathrm{H}^{\ell m}_{\mathrm{pw}}\right],\ \forall\mathbf{a}\in\mathbb{R}^p_*,\ \forall\ell<m,\ \ell,m\in\{1,\dots,k\}, \tag{3.3}$$
$$\left[\mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{1m} - \mathrm{H}^{1m}_{\mathrm{c}}, \mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{1m} + \mathrm{H}^{1m}_{\mathrm{c}}\right],\ \forall\mathbf{a}\in\mathbb{R}^p_*,\ \forall m\in\{2,\dots,k\}, \tag{3.4}$$

where

$$\mathrm{H}_{\mathrm{pw}}^{\ell m} = \|\mathbf{a}\|\sqrt{\frac{\mathrm{tr}(\mathbf{S}_\ell)}{n_\ell} + \frac{\mathrm{tr}(\mathbf{S}_m)}{n_m} + \widehat{\sigma}_{\ell m} h_{\mathrm{pw}}^{\ell m}}, \;\; \mathrm{H}_{\mathrm{c}}^{1m} = \|\mathbf{a}\|\sqrt{\frac{\mathrm{tr}(\mathbf{S}_1)}{n_1} + \frac{\mathrm{tr}(\mathbf{S}_m)}{n_m} + \widehat{\sigma}_{1m} h_{\mathrm{c}}^{1m}}.$$

In order to construct exact simultaneous confidence intervals (3.3) and (3.4), we need to find exact values $h_{\mathrm{pw}}^{\ell m}$ and h_{c}^{1m}. However, since it is difficult to find exact values $h_{\mathrm{pw}}^{\ell m}$ and h_{c}^{1m}, we give approximations for $h_{\mathrm{pw}}^{\ell m}$ and h_{c}^{1m} based on Bonferroni's inequality. Here, P_{pw} and P_{c} can be rewritten as follows.

$$P_{\mathrm{pw}} = 1 - \Pr\left[\bigcup_{1\le \ell < m \le k} \left\{H_{\ell m} \ge h_{\mathrm{pw}}^{\ell m}\right\}\right], \;\; P_{\mathrm{c}} = 1 - \Pr\left[\bigcup_{2\le m \le k} \left\{H_{1m} \ge h_{\mathrm{c}}^{1m}\right\}\right].$$

So, from Bonferroni's inequality, we obtain

$$P_{\mathrm{pw}} \ge 1 - \sum_{1\le \ell < m \le k} \Pr(H_{\ell m} \ge h_{\mathrm{pw}}^{\ell m}), \;\; P_{\mathrm{c}} \ge 1 - \sum_{2\le m \le k} \Pr(H_{1m} \ge h_{\mathrm{c}}^{1m}).$$

By using Lemma 2 and Theorem 1, we construct asymptotically conservative simultaneous confidence intervals by choosing $h_{\mathrm{pw}}^{\ell m}$ and h_{c}^{1m} so that $\Pr(H_{\ell m} \ge h_{\mathrm{pw}}^{\ell m}) = \alpha_{\mathrm{pw}} + o(1)$ and $\Pr(H_{1m} \ge h_{\mathrm{c}}^{1m}) = \alpha_{\mathrm{c}} + o(1)$. The specific forms of these simultaneous confidence intervals are obtained in the following way.

1. Simultaneous confidence intervals for pairwise multiple comparisons among mean vectors are given by

$$HCI_{\mathrm{pw1}} = \left[\mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{\ell m} - \mathrm{H}_{\mathrm{pw1}}^{\ell m}, \mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{\ell m} + \mathrm{H}_{\mathrm{pw1}}^{\ell m}\right], \;\forall \mathbf{a} \in \mathbb{R}_*^p, \;\forall \ell < m, \;\ell, m \in \{1, \dots, k\}, \tag{3.5}$$

where

$$\mathrm{H}_{\mathrm{pw1}}^{\ell m} = \|\mathbf{a}\|\sqrt{\frac{\mathrm{tr}(\mathbf{S}_\ell)}{n_\ell} + \frac{\mathrm{tr}(\mathbf{S}_m)}{n_m} + \widehat{\sigma}_{\ell m} z_{\alpha_{\mathrm{pw}}}}.$$

2. Simultaneous confidence intervals for multiple comparisons with a control among mean vectors are given by

$$HCI_{\mathrm{c1}} = \left[\mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{1m} - \mathrm{H}_{\mathrm{c1}}^{1m}, \mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{1m} + \mathrm{H}_{\mathrm{c1}}^{1m}\right], \;\forall \mathbf{a} \in \mathbb{R}_*^p, \;\forall m \in \{2, \dots, k\}, \tag{3.6}$$

where

$$\mathrm{H}_{\mathrm{c1}}^{1m} = \|\mathbf{a}\|\sqrt{\frac{\mathrm{tr}(\mathbf{S}_1)}{n_1} + \frac{\mathrm{tr}(\mathbf{S}_m)}{n_m} + \widehat{\sigma}_{1m} z_{\alpha_{\mathrm{c}}}}.$$

3. Simultaneous confidence intervals for pairwise multiple comparisons among mean vectors are given by

$$HCI_{\mathrm{pw2}} = \left[\mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{\ell m} - \mathrm{H}_{\mathrm{pw2}}^{\ell m}, \mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{\ell m} + \mathrm{H}_{\mathrm{pw2}}^{\ell m}\right], \;\forall \mathbf{a} \in \mathbb{R}_*^p, \;\forall \ell < m, \;\ell, m \in \{1, \dots, k\}, \tag{3.7}$$

where

$$\mathrm{H}_{\mathrm{pw2}}^{\ell m} = \|\mathbf{a}\|\sqrt{\frac{\mathrm{tr}(\mathbf{S}_\ell)}{n_\ell} + \frac{\mathrm{tr}(\mathbf{S}_m)}{n_m} + \widehat{\sigma}_{\ell m} \widehat{cf}_{\ell m}(z_{\alpha_{\mathrm{pw}}})}.$$

4. Simultaneous confidence intervals for multiple comparisons with a control among mean vectors are given by

$$HCI_{\mathrm{c2}} = \left[\mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{1m} - \mathrm{H}_{\mathrm{c2}}^{1m}, \mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{1m} + \mathrm{H}_{\mathrm{c2}}^{1m}\right], \;\forall \mathbf{a} \in \mathbb{R}_*^p, \;\forall m \in \{2, \dots, k\}, \tag{3.8}$$

where

$$\mathrm{H}_{\mathrm{c2}}^{1m} = \|\mathbf{a}\|\sqrt{\frac{\mathrm{tr}(\mathbf{S}_1)}{n_1} + \frac{\mathrm{tr}(\mathbf{S}_m)}{n_m} + \widehat{\sigma}_{1m} \widehat{cf}_{1m}(z_{\alpha_{\mathrm{c}}})}.$$

Simultaneous confidence intervals given by 1 and 2 are approximations using percentage points of the limit distribution of $H_{\ell m}$. Simultaneous confidence intervals given by 3 and 4 are approximations using the Cornish-Fisher expansion for $H_{\ell m}$. Also, we note that these four simultaneous confidence intervals (3.5)–(3.8) can be simply expressed when $k = 2$. See the following remark for details.

Remark 1. *If $k = 2$, simultaneous confidence intervals (3.5)–(3.8) are unified into the following two confidence intervals.*

$$HCI_1 = \left[\mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{12} - \mathrm{H}_1^{12},\ \mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{12} + \mathrm{H}_1^{12}\right],\ \forall\mathbf{a} \in \mathbb{R}_*^p,$$
$$HCI_2 = \left[\mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{12} - \mathrm{H}_2^{12},\ \mathbf{a}^\top\widehat{\boldsymbol{\delta}}_{12} + \mathrm{H}_2^{12}\right],\ \forall\mathbf{a} \in \mathbb{R}_*^p,$$

where

$$\mathrm{H}_1^{12} = \|\mathbf{a}\|\sqrt{\frac{\mathrm{tr}(\mathbf{S}_1)}{n_1} + \frac{\mathrm{tr}(\mathbf{S}_2)}{n_2} + \widehat{\sigma}_{12} z_\alpha}\ ,\ \mathrm{H}_2^{12} = \|\mathbf{a}\|\sqrt{\frac{\mathrm{tr}(\mathbf{S}_1)}{n_1} + \frac{\mathrm{tr}(\mathbf{S}_2)}{n_2} + \widehat{\sigma}_{12}\widehat{cf}_{12}(z_\alpha)}\ .$$

HCI_1 and HCI_2 are confidence intervals for the set of all linear combinations of two mean difference $\mathbf{a}^\top(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \mathbf{a}^\top\boldsymbol{\delta}_{12}$ for all $\mathbf{a} \in \mathbb{R}_*^p$.

With Lemma 1 and Theorem 1, we can obtain the following theorem. This theorem refers to convergence rates of the lower boundary of coverage probability of the proposed new confidence intervals.

Theorem 2. *The lower boundary of coverage probability for each simultaneous confidence intervals HCI_{pw1}, HCI_{c1}, HCI_{pw2}, and HCI_{c2} are defined as*

$$L_{\mathrm{pw1}} = 1 - \sum_{1\le\ell<m\le k} \Pr\left(H_{\ell m} \ge \widehat{\sigma}_{\ell m} z_{\alpha_{\mathrm{pw}}}\right),\ L_{\mathrm{c1}} = 1 - \sum_{2\le m\le k} \Pr\left(H_{1m} \ge \widehat{\sigma}_{1m} z_{\alpha_{\mathrm{c}}}\right),$$
$$L_{\mathrm{pw2}} = 1 - \sum_{1\le\ell<m\le k} \Pr\left\{H_{\ell m} \ge \widehat{\sigma}_{\ell m}\widehat{f}_{\ell m}(z_{\alpha_{\mathrm{pw}}})\right\},\ L_{\mathrm{c2}} = 1 - \sum_{2\le m\le k} \Pr\left\{H_{1m} \ge \widehat{\sigma}_{1m}\widehat{f}_{1m}(z_{\alpha_{\mathrm{c}}})\right\}.$$

Under (A1) *and* (A2)*, it holds that*

$$L_{\mathrm{pw1}} = 1 - \alpha + O(p^{\beta_{(1)}-1/2}),\ L_{\mathrm{c1}} = 1 - \alpha + O(p^{\beta_{(1)}-1/2}),$$
$$L_{\mathrm{pw2}} = 1 - \alpha + o(p^{\beta_{(1)}-1/2}),\ L_{\mathrm{c2}} = 1 - \alpha + o(p^{\beta_{(1)}-1/2}).$$

Proof. See, Appendix E. □

From this theorem, it can be confirmed that asymptotic conservatism is established for any proposed method. Also, we recommend the estimated Cornish-Fisher expansion-based simultaneous confidence intervals HCI_{pw2} and HCI_{c2} since L_{pw2} and L_{c2} converge toward nominal confidence $1 - \alpha$ faster than L_{pw1} and L_{c1}.

4. Empirical simulation studies

In this section, we perform Monte Carlo simulations with 10,000 trials in order to verify the superiority of proposed approximations and evaluate the accuracy of approximations in terms of coverage probability. Also, we show the robustness of proposed approximations under non-normality.

4.1. Empirical comparisons

In this section, we compare proposed simultaneous confidence intervals for pairwise comparisons HCI_{pw1} and HCI_{pw2} and for comparison with a control HCI_{c1} and HCI_{c2}, introduced in (3.5)–(3.8), with Takahashi et al. (2013)'s simultaneous confidence intervals for pairwise comparisons TCI_{pw1} and TCI_{pw2} and for comparison with a control TCI_{c1} and TCI_{c2} that were introduced in (2.1)–(2.4).

We calculate empirical coverage probabilities for these confidence intervals and compare them to nominal confidence levels $1 - \alpha$, $\alpha \in \{0.1, 0.05, 0.01\}$. Here, it is desirable that empirical coverage probabilities are equal to or higher than the nominal confidence level $1-\alpha$. In this simulation, we set the dimensions as $p \in \{100, 300, 500, 700\}$ and the sample sizes for each $k \in \{3, 5\}$ were set as follows.

$$\text{(I) } (n_1, n_2, n_3) \in \{(60, 60, 60), (40, 60, 80)\}$$
$$\text{(II) } (n_1, n_2, n_3, n_4, n_5) \in \{(60, 60, 60, 60, 60), (20, 40, 60, 80, 100)\}$$

We also set the covariance structures as follows.

$$\text{(I) } \boldsymbol{\Sigma}_1 = 5(0.5^{|l-m|}),\ \boldsymbol{\Sigma}_2 = 3(0.3^{|l-m|}),\ \boldsymbol{\Sigma}_3 = (0.1^{|l-m|})$$
$$\text{(II) } \boldsymbol{\Sigma}_1 = 5(0.5^{|l-m|}),\ \boldsymbol{\Sigma}_2 = 4(0.4^{|l-m|}),\ \boldsymbol{\Sigma}_3 = 3(0.3^{|l-m|}),\ \boldsymbol{\Sigma}_4 = 2(0.2^{|l-m|}),\ \boldsymbol{\Sigma}_5 = (0.1^{|l-m|})$$

(I) represents the setting at $k = 3$ and (II) represents the setting at $k = 5$.

Tables 1 and 2 summarize empirical coverage probabilities for each simultaneous confidence intervals for pairwise comparisons. In addition, Tables 3 and 4 summarize empirical coverage probabilities for each simultaneous confidence intervals for comparisons with a control.

First, we focus on a case of simultaneous confidence intervals for pairwise comparisons. From Tables 1 and 2, it can be seen that coverage probabilities of TCI_{pw1} and TCI_{pw2} are extremely smaller than nominal confidence level $1-\alpha$, even though coverage probabilities should be greater than or equal to $1-\alpha$. Therefore, Takahashi et al (2013)'s method is not recommended for use when homogeneity of variance-covariance matrices across groups is violated. On the other hand, it is obvious that proposed simultaneous confidence intervals HCI_{pw1} and HCI_{pw2} are close to nominal level $1-\alpha$. However, the normal approximation-based method HCI_{pw1} is often not conservative. It can be seen that coverage probability of the Cornish-Fisher expansion-based method HCI_{pw2} is close to nominal confidence level $1-\alpha$ and is often conservative.

The same consideration can be applied to the control case as for pairwise. In fact, the same tendency as in the case of pairwise comparisons can be confirmed from Tables 3 and 4. To summarize, we recommend the Cornish-Fisher expansion-based simultneous confidence intervals HCI_{pw2} and HCI_{c2}.

4.2. Robustness of the proposed approximation

In this subsection, we evaluate the robustness of the proposed simultaneous confidence intervals under non-normality in terms of coverage probability. We consider the following data generation model:

$$\mathbf{x}_{ij} = \boldsymbol{\Sigma}_i^{1/2}\mathbf{z}_{ij},\ i \in \{1, 2, 3\},\ j \in \{1, \ldots, n_i\},$$

where $\boldsymbol{\Sigma}_1 = 5(0.5^{|l-m|})$, $\boldsymbol{\Sigma}_2 = 3(0.3^{|l-m|})$, $\boldsymbol{\Sigma}_3 = (0.1^{|l-m|})$, $n_1 = n_2 = n_3 = 60$ and the random vector $\mathbf{z}_{ij} = \left(\mathrm{z}_{ijk}\right)$ has the following distributions:

$$\text{(D1)}\quad \mathbf{z}_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_p),$$
$$\text{(D2)}\quad \mathrm{z}_{ijk} = \mathrm{u}_{ijk}/\sqrt{5/4},\ \text{where } \mathrm{u}_{ijk} \overset{\text{iid}}{\sim} \mathcal{T}_{10},$$
$$\text{(D3)}\quad \mathbf{z}_{ij} = \sqrt{4/5}\mathbf{u}_{ij},\ \text{where } \mathbf{u}_{ij} \overset{\text{iid}}{\sim} \mathcal{T}(10, \mathbf{0}, \mathbf{I}_p),$$
$$\text{(D4)}\quad \mathrm{z}_{ijk} = \left(1 - \frac{9}{5\pi}\right)^{-1/2}\left(\mathrm{u}_{ijk} + \frac{3}{\sqrt{5\pi}}\right),\ \text{where } \mathrm{u}_{ijk} \overset{\text{iid}}{\sim} \mathcal{SN}(-3).$$

Here, $\mathcal{T}(10, \mathbf{0}, \mathbf{I}_p)$ denotes a multivariate t-distribution with degrees of freedom 10, location $\mathbf{0}$, and shape matrix $\mathbf{I}_p$. It should be noted that (D3) belongs to the class of elliptical distributions, whereas (D4) represents a case of asymmetric distribution.

Table 5 lists empirical coverage probabilities for HCI_{pw1}, HCI_{pw2}, HCI_{c1}, and HCI_{c2} under settings (D1)–(D4). The empirical coverage probabilities HCI_{pw2} and HCI_{c2} are larger than or equal to nominal confidence level 0.95 except under (D3). Alternatively, the empirical coverage probability under (D3) is extremely large compared to nominal confidence level 0.95. When assuming an elliptical population like (D3), there is concern that our proposed methods are not robust. To summarize, we expect that the proposed method is robust under non-normal settings such that each component of $\mathbf{z}_{ij}$ is independent, $\mathrm{E}(\mathrm{z}_{ijk}) = 0$, and $\mathrm{var}(\mathrm{z}_{ijk}) = 1$.

Table 1: This table summarizes empirical coverage probabilities for each simultaneous confidence intervals for pairwise comparisons. Row k specifies the number of groups, row n is where B stands for $(n_1, n_2, n_3) = (60, 60, 60)$ and UB stands for $(n_1, n_2, n_3) = (40, 60, 80)$; row p specifies the dimension, and row $1 - \alpha$ specifies the nominal confidence level. When the simultaneous confidence intervals are conservative (empirical coverage probabilities are greater than or equal to $1 - \alpha$), results are highlighted in bold.

k	n	p	$1-\alpha$	TCI_{pw1}	TCI_{pw2}	HCI_{pw1}	HCI_{pw2}
3	B	100	0.9	0.505	0.558	0.886	**0.914**
			0.95	0.596	0.665	0.931	**0.954**
			0.99	0.744	0.837	0.978	**0.990**
		300	0.9	0.159	0.180	**0.900**	**0.915**
			0.95	0.232	0.275	0.943	**0.955**
			0.99	0.417	0.508	0.983	**0.991**
		500	0.9	0.047	0.056	0.899	**0.914**
			0.95	0.087	0.108	0.945	**0.954**
			0.99	0.221	0.287	0.984	**0.990**
		700	0.9	0.016	0.019	**0.900**	**0.912**
			0.95	0.034	0.045	0.943	**0.952**
			0.99	0.116	0.155	0.985	**0.990**
	UB	100	0.9	0.093	0.113	0.895	**0.915**
			0.95	0.138	0.180	0.931	**0.953**
			0.99	0.254	0.361	0.976	**0.990**
		300	0.9	0.001	0.002	**0.900**	**0.916**
			0.95	0.004	0.005	0.941	**0.953**
			0.99	0.012	0.021	0.984	**0.991**
		500	0.9	0.000	0.000	**0.904**	**0.916**
			0.95	0.000	0.000	0.946	**0.957**
			0.99	0.000	0.001	0.984	**0.991**
		700	0.9	0.000	0.000	**0.906**	**0.914**
			0.95	0.000	0.000	0.946	**0.954**
			0.99	0.000	0.000	0.986	**0.991**

Table 2: This table summarizes empirical coverage probabilities for each simultaneous confidence intervals for pairwise comparisons. Row k specifies the number of groups, row n is where B stands for $(n_1, n_2, n_3, n_4, n_5) = (60, 60, 60, 60, 60)$, UB stands for $(n_1, n_2, n_3, n_4, n_5) = (20, 40, 60, 80, 100)$; row p specifies the dimension, and row $1 - \alpha$ specifies the nominal confidence level. When the simultaneous confidence intervals are conservative (empirical coverage probabilities are greater than or equal to $1 - \alpha$), results are highlighted in bold.

k	n	p	$1-\alpha$	TCI_{pw1}	TCI_{pw2}	HCI_{pw1}	HCI_{pw2}
5	B	100	0.9	0.223	0.295	0.871	**0.919**
			0.95	0.289	0.390	0.919	**0.956**
			0.99	0.438	0.596	0.971	**0.990**
		300	0.9	0.011	0.016	0.892	**0.919**
			0.95	0.020	0.031	0.933	**0.954**
			0.99	0.055	0.089	0.977	0.988
		500	0.9	0.000	0.001	0.896	**0.918**
			0.95	0.001	0.002	0.939	**0.958**
			0.99	0.004	0.008	0.983	**0.991**
		700	0.9	0.000	0.000	0.898	**0.917**
			0.95	0.000	0.000	0.943	**0.958**
			0.99	0.000	0.001	0.984	**0.990**
	UB	100	0.9	0.001	0.002	0.879	**0.918**
			0.95	0.002	0.004	0.920	**0.956**
			0.99	0.007	0.017	0.969	0.987
		300	0.9	0.000	0.000	**0.902**	**0.927**
			0.95	0.000	0.000	0.941	**0.960**
			0.99	0.000	0.000	0.981	**0.990**
		500	0.9	0.000	0.000	**0.909**	**0.927**
			0.95	0.000	0.000	0.946	**0.960**
			0.99	0.000	0.000	0.984	**0.991**
		700	0.9	0.000	0.000	**0.907**	**0.925**
			0.95	0.000	0.000	0.947	**0.963**
			0.99	0.000	0.000	0.985	**0.990**

Table 3: This table summarizes empirical coverage probabilities for each simultaneous confidence intervals for comparisons with a control. Row k specifies the number of groups, row n is where B stands for $(n_1, n_2, n_3) = (60, 60, 60)$, UB stands for $(n_1, n_2, n_3) = (40, 60, 80)$; row p specifies the dimension, and row $1-\alpha$ specifies the nominal confidence level. When the simultaneous confidence intervals are conservative (empirical coverage probabilities are greater than or equal to $1-\alpha$), results are highlighted in bold.

k	n	p	$1-\alpha$	TCI_{c1}	TCI_{c2}	HCI_{c1}	HCI_{c2}
3	B	100	0.9	0.457	0.490	0.903	**0.920**
			0.95	0.551	0.605	0.940	**0.956**
			0.99	0.719	0.808	0.979	**0.990**
		300	0.9	0.123	0.136	**0.904**	**0.914**
			0.95	0.194	0.222	0.944	**0.956**
			0.99	0.376	0.460	0.985	**0.991**
		500	0.9	0.034	0.038	**0.909**	**0.918**
			0.95	0.063	0.077	0.947	**0.956**
			0.99	0.184	0.235	0.985	**0.991**
		700	0.9	0.010	0.012	**0.907**	**0.914**
			0.95	0.023	0.027	0.948	**0.956**
			0.99	0.088	0.117	0.987	**0.991**
	UB	100	0.9	0.073	0.086	**0.907**	**0.922**
			0.95	0.112	0.147	0.942	**0.959**
			0.99	0.230	0.313	0.979	0.989
		300	0.9	0.000	0.000	**0.912**	**0.922**
			0.95	0.002	0.002	0.948	**0.960**
			0.99	0.008	0.013	0.985	**0.991**
		500	0.9	0.000	0.000	**0.912**	**0.921**
			0.95	0.000	0.000	**0.950**	**0.958**
			0.99	0.000	0.000	0.985	**0.991**
		700	0.9	0.000	0.000	**0.911**	**0.918**
			0.95	0.000	0.000	**0.950**	**0.956**
			0.99	0.000	0.000	0.988	**0.991**

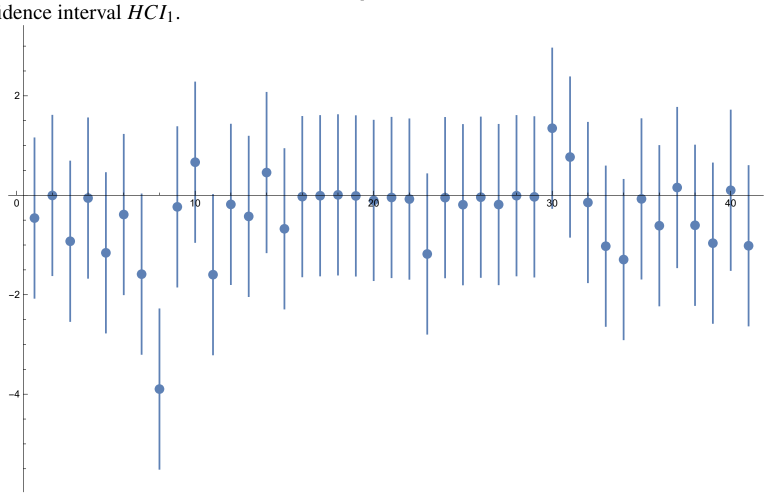
Table 4: This table summarizes empirical coverage probabilities for each simultaneous confidence intervals for comparisons with a control. Row k specifies the number of groups, row n is where B stands for $(n_1, n_2, n_3, n_4, n_5) = (60, 60, 60, 60, 60)$, UB stands for $(n_1, n_2, n_3, n_4, n_5) = (20, 40, 60, 80, 100)$; row p specifies the dimension, and row $1-\alpha$ specifies the nominal confidence level. When the simultaneous confidence intervals are conservative (empirical coverage probabilities are greater than or equal to $1-\alpha$), results are highlighted in bold.

k	n	p	$1-\alpha$	TCI_{c1}	TCI_{c2}	HCI_{c1}	HCI_{c2}
5	B	100	0.9	0.166	0.199	**0.901**	**0.927**
			0.95	0.226	0.286	0.936	**0.958**
			0.99	0.372	0.493	0.976	**0.991**
		300	0.9	0.006	0.007	**0.910**	**0.927**
			0.95	0.011	0.015	0.945	**0.962**
			0.99	0.036	0.055	0.982	**0.990**
		500	0.9	0.000	0.000	**0.916**	**0.930**
			0.95	0.001	0.001	**0.950**	**0.961**
			0.99	0.003	0.005	0.987	**0.992**
		700	0.9	0.000	0.000	**0.916**	**0.927**
			0.95	0.000	0.000	**0.952**	**0.962**
			0.99	0.000	0.000	0.986	**0.990**
	UB	100	0.9	0.001	0.002	**0.920**	**0.939**
			0.95	0.002	0.003	0.947	**0.965**
			0.99	0.008	0.016	0.979	**0.991**
		300	0.9	0.000	0.000	**0.932**	**0.943**
			0.95	0.000	0.000	**0.957**	**0.968**
			0.99	0.000	0.000	0.985	**0.990**
		500	0.9	0.000	0.000	**0.935**	**0.944**
			0.95	0.000	0.000	**0.962**	**0.971**
			0.99	0.000	0.000	0.989	**0.993**
		700	0.9	0.000	0.000	**0.933**	**0.942**
			0.95	0.000	0.000	**0.962**	**0.970**
			0.99	0.000	0.000	0.989	**0.993**

Table 5: This table shows empirical coverage probability for each distribution (D1)–(D4). We use the simulation settings $p = 500$ and $(n_1, n_2, n_3) = (60, 60, 60)$. The nominal confidence level is 0.95.

	(D1)		(D2)		(D3)		(D4)	
Pairwise	HCI_{pw1}	HCI_{pw2}	HCI_{pw1}	HCI_{pw2}	HCI_{pw1}	HCI_{pw2}	HCI_{pw1}	HCI_{pw2}
	0.943	**0.954**	0.944	**0.954**	**0.998**	**1.000**	0.941	**0.953**
Control	HCI_{c1}	HCI_{c2}	HCI_{c1}	HCI_{c2}	HCI_{c1}	HCI_{c2}	HCI_{c1}	HCI_{c2}
	0.950	**0.960**	0.947	**0.955**	**0.997**	**0.999**	0.947	**0.955**

Figure 1: The horizontal axis represents the number corresponding to each molecular descriptor and the vertical axis represents the measured value corresponding to each molecular descriptor. By setting $\mathbf{a}$ to $\mathbf{e}_1, \ldots, \mathbf{e}_p$ in HCI_1, we construct confidence intervals for each molecular descriptor. The solid lines denote confidence interval HCI_1.



4.3. An example of data analysis

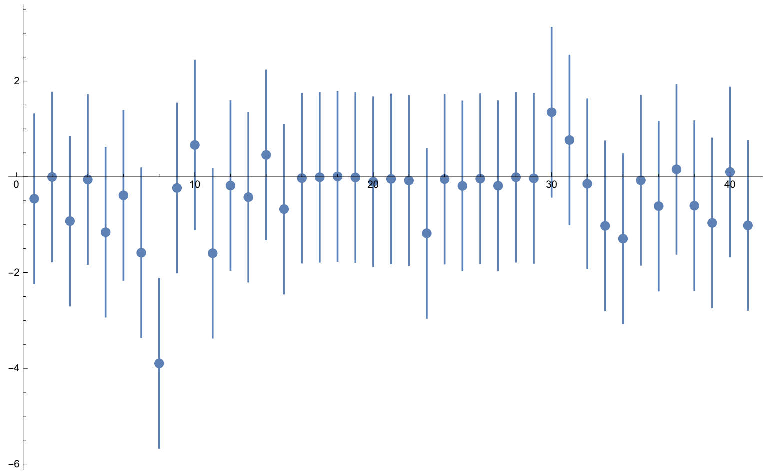
We apply our simultaneous confidence intervals to biodegradability experimental data. In Mansouri et al. (2013), this data set was used to develop QSAR (Quantitative Structure Activity Relationships) models for the study of relationships between chemical structure and biodegradation of molecules. Biodegradation experimental values of $n = 1055$ chemicals were collected from the webpage of the National Institute of Technology and Evaluation of Japan (NITE). These 1055 chemicals are divided into two groups by biodegradability, $n_1 = 356$ biodegradable organics (RB) and $n_2 = 699$ non-biodegradable organics (NRB). This data set contains $p = 41$ molecular descriptors for each chemical. Specifically, these $p = 41$ molecular descriptors are given by Mansouri et al. (2013).

By applying the test for equality of covariance matrices proposed by Li and Qin (2014), homogeneity of variance-covariance matrices across groups $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$ is rejected at a significance level of 0.05. Thus, using proposed confidence intervals HCI_{pw1} and HCI_{pw2}, that is, HCI_1 and HCI_2 stated in Remark 1, we construct simultaneous confidence intervals for each molecular descriptor without assuming that variance-covariance matrices are homogeneous across groups. Fig. 1 and Fig. 2 represent each simultaneous confidence intervals. As shown, there is a significant difference only in 8) Percentage of C atoms and no significant difference in other molecular descriptors. From this result, it is expected that Percentage of C atoms is related to the presence or absence of biodegradability.

5. Conclusions

In this study, we proposed a new procedure for pairwise multiple comparisons and multiple comparisons with a control among mean vectors under covariance heterogeneity in high-dimensional settings. Our main contribution is an extension of multiple comparisons procedures under assumption of homoscedasticity by Takahashi et al. (2013). For approximate interval estimation, we used the studentized L^2-type statistic, which is robust even if the assumption of homoscedasticity is not true. In order to derive a valid approximate simultaneous confidence intervals in high-dimensional settings, it is necessary to derive an accurate approximate percentage point for the studentized L^2-type statistic. A simple method is to use the normal approximation for studentized L^2-type statistic proposed by Chen and Qin (2010). However, it is empirically known that simultaneous confidence intervals applying normal approximation are often too loose or fail to capture the tail behavior of the resulting distribution (see, e.g., results of empirical studies by Nishiyama et al. (2013) and Hyodo et al. (2014)). Motivated by this point, we derived a Cornish-Fisher expansion of the studentized L^2-type statistic and proposed an approximate simultaneous confidence intervals. We also showed that the lower boundary of coverage probabilities of proposed new simultaneous confidence intervals convergence

Figure 2: The horizontal axis represents the number corresponding to each molecular descriptor and the vertical axis represents the measured value corresponding to each molecular descriptor. By setting $\mathbf{a}$ to $\mathbf{e}_1, \dots, \mathbf{e}_p$ in HCI_2, we construct confidence intervals for each molecular descriptor. The solid lines denote confidence interval HCI_2.



with the nominal confidence level; that is, our proposed simultaneous confidence intervals were asymptotically conservative. In addition, we compared empirical coverage probability of the two proposed simultaneous confidence intervals based on normal approximation and a Cornish-Fisher expansion with Takahashi et al (2013)'s simultaneous confidence intervals in numerical simulations, finding that Cornish-Fisher expansion-based confidence intervals are accurate and conservative, and perform better than others. Even in some non-normal settings, the proposed simultaneous confidence intervals generally work.

Appendix

A. Proof of Lemma 1

The mean and variance of $\widetilde{D}_{\ell m}$ are obtained by

$$\mathrm{E}(\widetilde{D}_{\ell m}) = \frac{n_m}{n_{\ell m}}\mathrm{tr}(\mathbf{\Sigma}_\ell) + \frac{n_\ell}{n_{\ell m}}\mathrm{tr}(\mathbf{\Sigma}_m),$$

$$\mathrm{var}(\widetilde{D}_{\ell m}) = 2\mathrm{tr}\left\{\left(\frac{n_m}{n_{\ell m}}\mathbf{\Sigma}_\ell + \frac{n_\ell}{n_{\ell m}}\mathbf{\Sigma}_m\right)^2\right\} = O(p).$$

Combining these results and from Markov inequality, under (A1) and (A2),

$$\widetilde{D}_{\ell m} / \sum_{i=1}^{k}(n_i - 1)/(n-k)\mathrm{tr}(\mathbf{\Sigma}_i) = m^*_{\ell m} + o_p(1). \tag{A. 1}$$

The mean and variance of $\mathrm{tr}(\mathbf{S})$ are obtained by

$$\mathrm{E}\{\mathrm{tr}(\mathbf{S})\} = \frac{1}{n-k}\sum_{i=1}^{k}(n_i - 1)\mathrm{tr}(\mathbf{\Sigma}_i),\ \ \mathrm{var}\{\mathrm{tr}(\mathbf{S})\} = \frac{2}{(n-k)^2}\sum_{i=1}^{k}(n_i - 1)\mathrm{tr}(\mathbf{\Sigma}_i^2) = O(1).$$

Combining these results and from Markov inequality, under (A1) and (A2),

$$\mathrm{tr}(\mathbf{S}) / \sum_{i=1}^{k}(n_i - 1)/(n-k)\mathrm{tr}(\mathbf{\Sigma}_i) = 1 + o_p(1). \tag{A. 2}$$

Combining (A. 1), (A. 2), and from Slutsky's theorem, $\widetilde{D}_{\ell m}/\mathrm{tr}(\mathbf{S}) = m^*_{\ell m} + o_p(1)$.

From (A. 2), under (A1) and (A2),

$$\widehat{\sigma} = \mathrm{tr}(\mathbf{S})\widehat{\sigma} / \sum_{i=1}^{k}(n_i - 1)/(n-k)\mathrm{tr}(\mathbf{\Sigma}_i)\{1 + o_p(1)\}. \tag{A. 3}$$

The expectation of $\widehat{\sigma}^2$ is given by

$$\begin{aligned}
\mathrm{E}[\{\mathrm{tr}(\mathbf{S})\}^2\widehat{\sigma}^2] =& \frac{2}{(n-k+2)(n-k-1)}\left[\sum_{i=1}^{k}\left\{(n_i-1) - \frac{(n_i-1)^2}{n-k}\right\}\{\mathrm{tr}(\mathbf{\Sigma}_i)\}^2\right. \\
&+ \sum_{i=1}^{k}\left\{(n_i-1)n_i - \frac{2(n_i-1)}{n-k}\right\}\mathrm{tr}(\mathbf{\Sigma}_i^2) + \sum_{i\neq j}^{k}(n_i-1)(n_j-1)\mathrm{tr}(\mathbf{\Sigma}_i\mathbf{\Sigma}_j) \\
&\left. - \sum_{i\neq j}^{k}\frac{(n_i-1)(n_j-1)}{n-k}\mathrm{tr}(\mathbf{\Sigma}_i)\mathrm{tr}(\mathbf{\Sigma}_j)\right] = O(p).
\end{aligned}$$

Combining these results and from Markov inequality, under (A1) and (A2),

$$\mathrm{tr}(\mathbf{S})\widehat{\sigma} = O_p(p^{1/2}) = o_p(p). \tag{A. 4}$$

Combining $\mathrm{tr}(\mathbf{\Sigma}_i) = O(p)$, $p = O\{\mathrm{tr}(\mathbf{\Sigma}_i)\}$, (A. 3), and (A. 4), $\widehat{\sigma} = o_p(1)$.

B. Proof of Lemma 2

We define the following mutually independent random variables:

$$\mathbf{z} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p),\ \mathbf{Z}_1 \sim \mathcal{MN}_{p,n_i-1}(\mathbf{O}, \mathbf{I}_p, \mathbf{I}_{n_i-1}),\ \mathbf{Z}_2 \sim \mathcal{MN}_{p,n_j-1}(\mathbf{O}, \mathbf{I}_p, \mathbf{I}_{n_j-1}).$$

Using $\mathbf{z}$, $\mathbf{Z}_1$, and $\mathbf{Z}_2$, we can write H_{ij} as

$$n_{\ell m}\left\{\|\widehat{\boldsymbol{\delta}}_{\ell m} - \boldsymbol{\delta}_{\ell m}\|^2 - \frac{\mathrm{tr}(\mathbf{S}_\ell)}{n_\ell} - \frac{\mathrm{tr}(\mathbf{S}_m)}{n_m}\right\} \overset{\mathcal{L}}{=} Y, \tag{A. 5}$$

where

$$Y = \mathbf{z}^\top\mathbf{\Lambda}_{\ell m}\mathbf{z} - \frac{n_{\ell m}\mathrm{tr}(\mathbf{\Sigma}_\ell\mathbf{Z}_1\mathbf{Z}_1^\top)}{(n_\ell-1)n_\ell} - \frac{n_{\ell m}\mathrm{tr}(\mathbf{\Sigma}_m\mathbf{Z}_2\mathbf{Z}_2^\top)}{(n_m-1)n_m}.$$

Here, $A \overset{\mathcal{L}}{=} B$ means that A and B have the same distribution and

$$\mathbf{\Lambda} = \frac{n_{\ell m}}{n_\ell}\mathbf{\Sigma}_\ell + \frac{n_{\ell m}}{n_m}\mathbf{\Sigma}_m.$$

Let $\tilde{\sigma}_{\ell m} = n_{\ell m}\sigma_{\ell m}$. From $\mathrm{E}(n_{\ell m}^2\widehat{\sigma}_{\ell m}^2 - \tilde{\sigma}_{\ell m}^2) = 0$ and $\mathrm{E}\{(n_{\ell m}^2\widehat{\sigma}_{\ell m}^2 - \tilde{\sigma}_{\ell m}^2)^2\} = O(n_{\ell m}^{-1}p^{1+2\beta_{(1)}})$,

$$E_{\ell m} = \frac{n_{\ell m}^2\widehat{\sigma}_{\ell m}^2 - \tilde{\sigma}_{\ell m}^2}{\tilde{\sigma}_{\ell m}^2} = O_p(n_{\ell m}^{-1/2}p^{\beta_{(1)}-1/2}). \tag{A. 6}$$

Combining (A. 5) and (A. 6),

$$H_{\ell m} \overset{\mathcal{L}}{=} \frac{Y}{\widetilde{\sigma}_{\ell m}}(1 + E_{\ell m})^{-1/2} = \frac{Y}{\tilde{\sigma}_{\ell m}}(1 - E_{\ell m}/2) + o_p(n_{\ell m}^{-1/2}p^{\beta_{(1)}-1/2}). \tag{A. 7}$$

Using the expression (A. 7), the first to third moments can be evaluated as following:

$$
\begin{aligned}
\mathrm{E}\left(H_{\ell m}\right) =& 8\widetilde{\sigma}_{\ell m}^{-3}\left\{\sum_{g\in\{\ell,m\}}\frac{n_{\ell m}^{3}\mathrm{tr}(\mathbf{\Sigma}_g^3)}{n_g^2(n_g-1)^2}+\frac{n_{\ell m}^{3}\mathrm{tr}(\mathbf{\Sigma}_\ell^2\mathbf{\Sigma}_m)}{(n_\ell-1)n_\ell^2 n_m}+\frac{n_{\ell m}^{3}\mathrm{tr}(\mathbf{\Sigma}_m^2\mathbf{\Sigma}_\ell)}{(n_j-1)n_m^2 n_\ell}\right\}+o(n_{\ell m}^{-1/2}p^{\beta_{(1)}-1/2}) \\
=& O(n_{\ell m}^{-1}p^{\beta_{(1)}-1/2})+o(n_{\ell m}^{-1/2}p^{\beta_{(1)}-1/2})=o(p^{\beta_{(1)}-1/2}), && \text{(A. 8)}\\
\mathrm{E}\left(H_{\ell m}^2\right) =& 1-32\widetilde{\sigma}_{\ell m}^{-4}\left\{\sum_{g\in\{\ell,m\}}\frac{3n_{\ell m}^{4}\mathrm{tr}(\mathbf{\Sigma}_g^4)}{n_g^3(n_g-1)^3}+\frac{2n_{\ell m}^{4}\mathrm{tr}(\mathbf{\Sigma}_\ell^3\mathbf{\Sigma}_m)}{(n_\ell-1)^2 n_\ell^3 n_m}+\frac{2n_{\ell m}^{4}\mathrm{tr}(\mathbf{\Sigma}_\ell^2\mathbf{\Sigma}_m^2)}{(n_\ell-1)n_\ell^2(n_m-1)n_m^2}\right.\\
&\left.+\frac{2n_{\ell m}^{4}\mathrm{tr}(\mathbf{\Sigma}_m^3\mathbf{\Sigma}_\ell)}{(n_m-1)^2 n_m^3 n_\ell}\right\}+o(n_{\ell m}^{-1/2}p^{\beta_{(1)}-1/2})\\
=& 1+O(n_{\ell m}^{-2}p^{2\beta_{(1)}-1})+o(n_{\ell m}^{-1/2}p^{\beta_{(1)}-1/2})=1+o(p^{\beta_{(1)}-1/2}), && \text{(A. 9)}\\
\mathrm{E}\left(H_{\ell m}^3\right) =& 8\widetilde{\sigma}_{\ell m}^{-3}b_{\ell m}+O(n_{\ell m}^{-1}p^{\beta_{(1)}-1/2})+o(n_{\ell m}^{-1/2}p^{\beta_{(1)}-1/2})=8\widetilde{\sigma}_{\ell m}^{-3}b_{\ell m}+o(p^{\beta_{(1)}-1/2}). && \text{(A. 10)}
\end{aligned}
$$

From (A. 8)–(A. 10), the first to third cumulants are obtained as

$$
\kappa_1\left(H_{\ell m}\right)=o(p^{\beta_{(1)}-1/2}),\ \kappa_2\left(H_{\ell m}\right)=1+o(p^{\beta_{(1)}-1/2}),\ \kappa_3\left(H_{\ell m}\right)=\frac{8b_{\ell m}}{\widetilde{\sigma}_{\ell m}}+o(p^{\beta_{(1)}-1/2}).
$$

Thus the Edgeworth expansion of $H_{\ell m}$ is given by (3.1).

C. Proof of Lemma 3

The estimator $\widehat{\sigma}_{\ell m}^2$ can be expressed as $\widehat{\sigma}_{\ell m}^2 = A_1 + A_2 + A_3$, where

$$
\begin{aligned}
A_1 =& \frac{2(n_\ell-1)}{n_\ell(n_\ell+1)(n_\ell-2)}\left[\mathrm{tr}(\mathbf{S}_\ell^2)-\frac{\{\mathrm{tr}(\mathbf{S}_\ell)\}^2}{n_\ell-1}\right],\\
A_2 =& \frac{2(n_m-1)}{n_m(n_m+1)(n_m-2)}\left[\mathrm{tr}(\mathbf{S}_m^2)-\frac{\{\mathrm{tr}(\mathbf{S}_m)\}^2}{n_m-1}\right],\\
A_3 =& \frac{4\mathrm{tr}(\mathbf{S}_\ell\mathbf{S}_m)}{n_\ell n_m}.
\end{aligned}
$$

The expectation of each term in $\widehat{\sigma}_{\ell m}^2$ can be evaluated as follows.

$$
\mathrm{E}(A_1)=\frac{2\mathrm{tr}(\mathbf{\Sigma}_\ell^2)}{n_\ell(n_\ell-1)},\ \mathrm{E}(A_2)=\frac{2\mathrm{tr}(\mathbf{\Sigma}_m^2)}{n_m(n_m-1)},\ \mathrm{E}(A_3)=\frac{4\mathrm{tr}(\mathbf{\Sigma}_\ell\mathbf{\Sigma}_m)}{n_\ell n_m}.
$$

Thus $\mathrm{E}(\widehat{\sigma}^2_{\ell m}) = \sigma^2_{\ell m}$. The variance of each term in $\widehat{\sigma}^2_{\ell m}$ can be evaluated as follows.

$$\begin{aligned}
\mathrm{var}(A_1) =&16\left[\frac{\{\mathrm{tr}(\mathbf{\Sigma}_\ell^2)\}^2}{(n_\ell-2)(n_\ell-1)^2n_\ell^2(n_\ell+1)} + \frac{\left(2n_\ell^2-n_\ell-7\right)\mathrm{tr}(\mathbf{\Sigma}_\ell^4)}{(n_\ell-2)(n_\ell-1)^3n_\ell^2(n_\ell+1)}\right]\\
=&O\left\{\frac{(p^{2\beta_{(1)}}+p)^2}{p^6}\right\}+O\left(\frac{p^{4\beta_{(1)}}+p}{p^5}\right) = O(p^{4\beta_{(1)}-5})+O(p^{-4}), \qquad \text{(A. 11)}\\
\mathrm{var}(A_2) =&16\left[\frac{\{\mathrm{tr}(\mathbf{\Sigma}_m^2)\}^2}{(n_m-2)(n_m-1)^2n_m^2(n_m+1)} + \frac{\left(2n_m^2-n_m-7\right)\mathrm{tr}(\mathbf{\Sigma}_m^4)}{(n_m-2)(n_m-1)^3n_m^2(n_m+1)}\right]\\
=&O\left\{\frac{(p^{2\beta_{(1)}}+p)^2}{p^6}\right\}+O\left(\frac{p^{4\beta_{(1)}}+p}{p^5}\right) = O(p^{4\beta_{(1)}-5})+O(p^{-4}), \qquad \text{(A. 12)}\\
\mathrm{var}(A_3) =&16\left[\frac{\{\mathrm{tr}(\mathbf{\Sigma}_\ell\mathbf{\Sigma}_m)\}^2}{(n_\ell-1)n_\ell^2(n_m-1)n_m^2} + \frac{\mathrm{tr}(\mathbf{\Sigma}_\ell^2)\mathrm{tr}(\mathbf{\Sigma}_m^2)}{(n_\ell-1)n_\ell^2(n_m-1)n_m^2}\right.\\
&\left.+\frac{n_{\ell m}\mathrm{tr}(\mathbf{\Sigma}_\ell^2\mathbf{\Sigma}_m^2)}{(n_\ell-1)n_\ell^2(n_m-1)n_m^2} + \frac{(n_{\ell m}-2)\mathrm{tr}\{(\mathbf{\Sigma}_\ell\mathbf{\Sigma}_m)^2\}}{(n_\ell-1)n_\ell^2(n_m-1)n_m^2}\right]\\
=&O\left\{\frac{(p^{2\beta_{(1)}}+p)^2}{p^6}\right\}+O\left(\frac{p^{4\beta_{(1)}}+p}{p^5}\right) = O(p^{4\beta_{(1)}-5})+O(p^{-4}). \qquad \text{(A. 13)}
\end{aligned}$$

From Hölder's inequality,

$$(\widehat{\sigma}^2_{\ell m}-\sigma^2_{\ell m})^2 = \left[\sum_{i=1}^3\{A_i-\mathrm{E}(A_i)\}\right]^2 \le 3\sum_{i=1}^3\{A_i-\mathrm{E}(A_i)\}^2.$$

By combining this result with (A. 11)–(A. 13), we have

$$\mathrm{var}(\widehat{\sigma}^2_{\ell m}) \le 3\sum_{i=1}^3\mathrm{var}(A_i) = O(p^{4\beta_{(1)}-5})+O(p^{-4}). \qquad \text{(A. 14)}$$

Note that $\sigma^2 = O(p^{-1})$ under (A1) and (A2). Thus $\mathrm{var}(\widehat{\sigma}^2_{\ell m})/\sigma^4 = O(p^{4\beta_{(1)}-3})+O(p^{-2}) = \mathrm{o}(1)$. That is, the ratio consistency of $\widehat{\sigma}^2_{\ell m}$ is shown.

The estimator $\widehat{b}^2_{\ell m}$ can be expressed as $\widehat{\sigma}^2_{\ell m} = B_1+B_2+B_3+B_4$, where

$$\begin{aligned}
B_1 =&\frac{(n_\ell-1)^2}{(n_\ell-3)n_\ell^2(n_\ell+1)(n_\ell+3)}\left[\mathrm{tr}(\mathbf{S}_\ell^3)-\frac{3\mathrm{tr}(\mathbf{S}_\ell^2)\mathrm{tr}(\mathbf{S}_\ell)}{n_\ell-1}+\frac{2\{\mathrm{tr}(\mathbf{S}_\ell)\}^3}{(n_\ell-1)^2}\right],\\
B_2 =&\frac{(n_m-1)^2}{(n_m-3)n_m^2(n_m+1)(n_m+3)}\left[\mathrm{tr}(\mathbf{S}_m^3)-\frac{3\mathrm{tr}(\mathbf{S}_m^2)\mathrm{tr}(\mathbf{S}_m)}{n_m-1}+\frac{2\{\mathrm{tr}(\mathbf{S}_m)\}^3}{(n_m-1)^2}\right],\\
B_3 =&\frac{3(n_\ell-1)^2}{(n_\ell-2)n_\ell^2(n_\ell+1)n_m}\left\{\mathrm{tr}(\mathbf{S}_\ell^2\mathbf{S}_m)-\frac{\mathrm{tr}(\mathbf{S}_\ell\mathbf{S}_m)\mathrm{tr}(\mathbf{S}_\ell)}{n_\ell-1}\right\},\\
B_4 =&\frac{3(n_m-1)^2}{(n_m-2)n_m^2(n_m+1)n_\ell}\left\{\mathrm{tr}(\mathbf{S}_m^2\mathbf{S}_\ell)-\frac{\mathrm{tr}(\mathbf{S}_m\mathbf{S}_\ell)\mathrm{tr}(\mathbf{S}_m)}{n_m-1}\right\}.
\end{aligned}$$

The expectation of each term in $\widehat{\sigma}^2_{\ell m}$ can be evaluated as follows.

$$\mathrm{E}(B_1) = \frac{(n_\ell-2)\mathrm{tr}(\mathbf{\Sigma}_\ell^3)}{n_\ell^2(n_{\ell m}-1)^2},\ \mathrm{E}(B_2) = \frac{(n_m-2)\mathrm{tr}(\mathbf{\Sigma}_m^3)}{n_m^2(n_m-1)^2},$$
$$\mathrm{E}(B_3) = \frac{3\mathrm{tr}(\mathbf{\Sigma}_\ell^2\mathbf{\Sigma}_m)}{n_\ell^2n_m},\ \mathrm{E}(B_4) = \frac{3\mathrm{tr}(\mathbf{\Sigma}_\ell\mathbf{\Sigma}_m^2)}{n_\ell n_m^2}.$$

Thus $\mathrm{E}(\widehat{b_{\ell m}}) = b_{\ell m}$. The variance of each term in $b_{\ell m}$ can be evaluated as follows.

$$\begin{aligned}
\mathrm{E}(B_1) =& \frac{6(n_\ell - 2)\{\mathrm{tr}(\mathbf{\Sigma}_\ell^2)\}^3}{(n_\ell - 3)(n_\ell - 1)^3 n_\ell^4 (n_\ell + 1)(n_\ell + 3)} + \frac{18(n_\ell - 2)\{\mathrm{tr}(\mathbf{\Sigma}_\ell^3)\}^2}{(n_\ell - 1)^4 n_\ell^4 (n_\ell + 1)} \\
&+ \frac{18(n_\ell - 2)\left(n_\ell^2 + n_\ell - 14\right)\mathrm{tr}(\mathbf{\Sigma}_\ell^4)\mathrm{tr}(\mathbf{\Sigma}_\ell^2)}{(n_\ell - 3)(n_\ell - 1)^4 n_\ell^4 (n_\ell + 1)(n_\ell + 3)} \\
&+ \frac{6(n_\ell - 2)\left(3n_\ell^4 + 3n_\ell^3 - 47n_\ell^2 - 47n_\ell + 248\right)\mathrm{tr}(\mathbf{\Sigma}_\ell^6)}{(n_\ell - 3)(n_\ell - 1)^5 n_\ell^4 (n_\ell + 1)(n_\ell + 3)} \\
=& O(p^{6\beta_{(1)} - 7}) + O(p^{-6}), \qquad \text{(A. 15)}
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}(B_2) =& \frac{6(n_m - 2)\{\mathrm{tr}(\mathbf{\Sigma}_m^2)\}^3}{(n_m - 3)(n_m - 1)^3 n_m^4 (n_m + 1)(n_m + 3)} + \frac{18(n_m - 2)\{\mathrm{tr}(\mathbf{\Sigma}_m^3)\}^2}{(n_m - 1)^4 n_m^4 (n_m + 1)} \\
&+ \frac{18(n_m - 2)\left(n_m^2 + n_m - 14\right)\mathrm{tr}(\mathbf{\Sigma}_m^4)\mathrm{tr}(\mathbf{\Sigma}_m^2)}{(n_m - 3)(n_m - 1)^4 n_m^4 (n_m + 1)(n_m + 3)} \\
&+ \frac{6(n_m - 2)\left(3n_m^4 + 3n_m^3 - 47n_m^2 - 47n_m + 248\right)\mathrm{tr}(\mathbf{\Sigma}_m^6)}{(n_m - 3)(n_m - 1)^5 n_m^4 (n_m + 1)(n_m + 3)} \\
=& O(p^{6\beta_{(1)} - 7}) + O(p^{-6}), \qquad \text{(A. 16)}
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}(B_3) =& \frac{9\{\mathrm{tr}(\mathbf{\Sigma}_\ell^2)\}^2\mathrm{tr}(\mathbf{\Sigma}_m^2)}{(n_\ell - 2)n_\ell^4(n_\ell + 1)(n_m - 1)n_m^2} + \frac{9\mathrm{tr}(\mathbf{\Sigma}_\ell^2)\{\mathrm{tr}(\mathbf{\Sigma}_\ell\mathbf{\Sigma}_m)\}^2}{(n_\ell - 2)n_\ell^4(n_\ell + 1)(n_m - 1)n_m^2} \\
&+ \frac{18\mathrm{tr}(\mathbf{\Sigma}_\ell^3)\mathrm{tr}(\mathbf{\Sigma}_\ell\mathbf{\Sigma}_m^2)}{(n_\ell - 1)n_\ell^4(n_m - 1)n_m^2} + \frac{18\left(n_\ell^2 + n_\ell n_m - 2n_\ell - n_m - 1\right)\{\mathrm{tr}(\mathbf{\Sigma}_\ell^2\mathbf{\Sigma}_m)\}^2}{(n_\ell - 2)(n_\ell - 1)n_\ell^4(n_\ell + 1)(n_m - 1)n_m^2} \\
&+ \frac{9\mathrm{tr}(\mathbf{\Sigma}_\ell^2)\mathrm{tr}\{(\mathbf{\Sigma}_\ell\mathbf{\Sigma}_m)^2\}}{(n_\ell - 2)n_\ell^4(n_\ell + 1)n_m^2} + \frac{9\left(n_\ell^2 + n_\ell n_m + n_\ell - n_m - 6\right)\mathrm{tr}(\mathbf{\Sigma}_\ell^2)\mathrm{tr}(\mathbf{\Sigma}_\ell^2\mathbf{\Sigma}_m^2)}{(n_\ell - 2)(n_\ell - 1)n_\ell^4(n_\ell + 1)(n_m - 1)n_m^2} \\
&+ \frac{9\left(n_\ell^2 - 5\right)\mathrm{tr}(\mathbf{\Sigma}_\ell^4)\mathrm{tr}(\mathbf{\Sigma}_m^2)}{(n_\ell - 2)(n_\ell - 1)n_\ell^4(n_\ell + 1)(n_m - 1)n_m^2}
\end{aligned}$$

$$
\begin{aligned}
&+\frac{18\left(n_\ell^2-5\right)\operatorname{tr}(\boldsymbol{\Sigma}_\ell\boldsymbol{\Sigma}_m)\operatorname{tr}(\boldsymbol{\Sigma}_\ell^3\boldsymbol{\Sigma}_m)}{(n_\ell-2)(n_\ell-1)n_\ell^4(n_\ell+1)(n_m-1)n_m^2}\\
&+\frac{18\left(2n_\ell^2n_m-n_\ell^2-n_\ell n_m-7n_m+5\right)\operatorname{tr}(\boldsymbol{\Sigma}_\ell^3\boldsymbol{\Sigma}_m\boldsymbol{\Sigma}_\ell\boldsymbol{\Sigma}_m)}{(n_\ell-2)(n_\ell-1)n_\ell^4(n_\ell+1)(n_m-1)n_m^2}\\
&+\frac{9\left(n_\ell^3+n_\ell^2n_m+2n_\ell^2-n_\ell-5n_m-14\right)\operatorname{tr}(\boldsymbol{\Sigma}_\ell^4\boldsymbol{\Sigma}_m^2)}{(n_\ell-2)(n_\ell-1)n_\ell^4(n_\ell+1)(n_m-1)n_m^2}\\
&+\frac{9\left(n_\ell^3+3n_\ell^2n_m-4n_\ell^2-2n_\ell n_m+3n_\ell-9n_m+4\right)\operatorname{tr}\{(\boldsymbol{\Sigma}_\ell^2\boldsymbol{\Sigma}_m)^2\}}{(n_\ell-2)(n_\ell-1)n_\ell^4(n_\ell+1)(n_m-1)n_m^2}\\
&=O(p^{6\beta_{(1)}-7})+O(p^{-6}), \qquad \text{(A. 17)}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}(B_4)=&\frac{9\operatorname{tr}(\boldsymbol{\Sigma}_m^2)\{\operatorname{tr}(\boldsymbol{\Sigma}_\ell\boldsymbol{\Sigma}_m)\}^2}{(n_\ell-1)n_\ell^2(n_m-2)n_m^4(n_m+1)}+\frac{9\operatorname{tr}(\boldsymbol{\Sigma}_\ell^2)\{\operatorname{tr}(\boldsymbol{\Sigma}_\ell\boldsymbol{\Sigma}_m)\}^2}{(n_\ell-1)n_\ell^2(n_m-2)n_m^4(n_m+1)}\\
&+\frac{18\operatorname{tr}(\boldsymbol{\Sigma}_m^3)\operatorname{tr}(\boldsymbol{\Sigma}_\ell^2\boldsymbol{\Sigma}_m)}{(n_\ell-1)n_\ell^2(n_m-1)n_m^4}+\frac{18\left(n_\ell n_m-n_\ell+n_m^2-2n_m-1\right)\{\operatorname{tr}(\boldsymbol{\Sigma}_\ell\boldsymbol{\Sigma}_m^2)\}^2}{(n_\ell-1)n_\ell^2(n_m-2)(n_m-1)n_m^4(n_m+1)}\\
&+\frac{9\operatorname{tr}(\boldsymbol{\Sigma}_m^2)\operatorname{tr}\{(\boldsymbol{\Sigma}_\ell\boldsymbol{\Sigma}_m)^2\}}{n_\ell^2(n_m-2)n_m^4(n_m+1)}+\frac{18\left(n_m^2-5\right)\operatorname{tr}(\boldsymbol{\Sigma}_\ell\boldsymbol{\Sigma}_m^3)\operatorname{tr}(\boldsymbol{\Sigma}_\ell\boldsymbol{\Sigma}_m)}{(n_\ell-1)n_\ell^2(n_m-2)(n_m-1)n_m^4(n_m+1)}\\
&+\frac{9\left(n_\ell n_m-n_\ell+n_m^2+n_m-6\right)\operatorname{tr}(\boldsymbol{\Sigma}_m^2)\operatorname{tr}(\boldsymbol{\Sigma}_\ell^2\boldsymbol{\Sigma}_m^2)}{(n_\ell-1)n_\ell^2(n_m-2)(n_m-1)n_m^4(n_m+1)}\\
&+\frac{9\left(n_m^2-5\right)\operatorname{tr}(\boldsymbol{\Sigma}_\ell^2)\operatorname{tr}(\boldsymbol{\Sigma}_m^4)}{(n_\ell-1)n_\ell^2(n_m-2)(n_m-1)n_m^4(n_m+1)}\\
&+\frac{18\left(2n_\ell n_m^2-n_\ell n_m-7n_\ell-n_m^2+5\right)\operatorname{tr}(\boldsymbol{\Sigma}_\ell\boldsymbol{\Sigma}_m\boldsymbol{\Sigma}_\ell\boldsymbol{\Sigma}_m^3)}{(n_\ell-1)n_\ell^2(n_m-2)(n_m-1)n_m^4(n_m+1)}\\
&+\frac{9\left(n_\ell n_m^2-5n_\ell+n_m^3+2n_m^2-n_m-14\right)\operatorname{tr}(\boldsymbol{\Sigma}_\ell^2\boldsymbol{\Sigma}_m^4)}{(n_\ell-1)n_\ell^2(n_m-2)(n_m-1)n_m^4(n_m+1)}\\
&+\frac{9\left(3n_\ell n_m^2-2n_\ell n_m-9n_\ell+n_m^3-4n_m^2+3n_m+4\right)\operatorname{tr}\{(\boldsymbol{\Sigma}_\ell\boldsymbol{\Sigma}_m^2)^2\}}{(n_\ell-1)n_\ell^2(n_m-2)(n_m-1)n_m^4(n_m+1)}\\
=&O(p^{6\beta_{(1)}-7})+O(p^{-6}). \qquad \text{(A. 18)}
\end{aligned}
$$

From Hölder's inequality,

$$
(\widehat{b}_{\ell m}-b_{\ell m})^2=\left[\sum_{i=1}^4\{B_i-\mathrm{E}(B_i)\}\right]^2\le 4\sum_{i=1}^4\{B_i-\mathrm{E}(B_i)\}^2.
$$

By combining this result with (A. 15)–(A. 18), we have

$$
\operatorname{var}(\widehat{b}_{\ell m})\le 4\sum_{i=1}^4\operatorname{var}(B_i)=O(p^{6\beta_{(1)}-7})+O(p^{-6}). \qquad \text{(A. 19)}
$$

Note that $b_{\ell m}=O(p^{-2}+p^{3\beta_{(1)}-3})$ under (A1) and (A2). Thus $\operatorname{var}(\widehat{b}_{\ell m})/b_{\ell m}^2=O(p^{6(\beta_{(1)}-1/2)})+O(p^{-2})=\mathrm{o}(1)$. That is, the ratio consistency of $\widehat{b}_{\ell m}$ is shown.

D. Proof of Theorem 1

This proof is proved in the same manner of Theorem 1 in Hall (1983). For any $x \in [-\lambda, \lambda]$ and any $\varepsilon > 0$,

$$\begin{aligned}\Pr\left\{H_{\ell m} \le \widehat{f_{\ell m}}(x)\right\} \le& \Pr\left\{H_{\ell m} \le \widehat{f_{\ell m}}(x),\ |\widehat{b}_{\ell m}/\sigma^3_{\ell m} - b_{\ell m}/\sigma^3_{\ell m}| < \varepsilon,\ |\widehat{\sigma}^2_{\ell m}/\sigma^2_{\ell m} - 1| < \varepsilon\right\} \\ &+ \Pr\left(|\widehat{b}_{\ell m}/\sigma^3_{\ell m} - b_{\ell m}/\sigma^3_{\ell m}| \ge \varepsilon\right) + \Pr\left(|\widehat{\sigma}^2_{\ell m}/\sigma^2_{\ell m} - 1| \ge \varepsilon\right). \end{aligned} \tag{A. 20}$$

If $|\widehat{b}_{\ell m}/\sigma^3_{\ell m} - b_{\ell m}/\sigma^3_{\ell m}| < \varepsilon$ and $|\widehat{\sigma}^2_{\ell m}/\sigma^2_{\ell m} - 1| < \varepsilon$ hold,

$$\left|\frac{\widehat{b}_{\ell m}}{\widehat{\sigma}^3_{\ell m}} - \frac{b_{\ell m}}{\sigma^3_{\ell m}}\right| < \eta(\varepsilon) = \max\left\{\left|\frac{b - \sigma^3\varepsilon}{\sigma^3(1+\varepsilon)^{3/2}} - \frac{b}{\sigma^3}\right|,\ \left|\frac{b + \sigma^3\varepsilon}{\sigma^3(1-\varepsilon)^{3/2}} - \frac{b}{\sigma^3}\right|\right\}. \tag{A. 21}$$

Combining (A. 20) and (A. 21), for any $x \in [-\lambda, \lambda]$ and any $\varepsilon > 0$,

$$\begin{aligned}\Pr\left\{H_{\ell m} \le \widehat{f_{\ell m}}(x)\right\} \le& \Pr\left\{H_{\ell m} \le f_{\ell m}(x) + (4/3)\eta(\varepsilon)(\lambda^2+1)\right\} \\ &+ \Pr\left(|\widehat{b}_{\ell m}/\sigma^3_{\ell m} - b_{\ell m}/\sigma^3_{\ell m}| \ge \varepsilon\right) + \Pr\left(|\widehat{\sigma}^2_{\ell m}/\sigma^2_{\ell m} - 1| \ge \varepsilon\right). \end{aligned} \tag{A. 22}$$

From Markov inequality, (A. 14), and (A. 19), under (A1) and (A2),

$$\Pr\left(\left|\frac{\widehat{b}_{\ell m}}{\sigma^3_{\ell m}} - \frac{b_{\ell m}}{\sigma^3_{\ell m}}\right| \ge \varepsilon\right) \le \frac{\mathrm{Var}(\widehat{b}_{\ell m})}{\varepsilon^2\sigma^6_{\ell m}} = O(\varepsilon^{-2}p^{-3}) + O(\varepsilon^{-2}p^{-4+6\beta_{(1)}}), \tag{A. 23}$$

$$\Pr\left(\left|\frac{\widehat{\sigma}^2_{\ell m}}{\sigma^2_{\ell m}} - 1\right| \ge \varepsilon\right) \le \frac{\mathrm{Var}(\widehat{\sigma}^2_{\ell m})}{\varepsilon^2\sigma^4_{\ell m}} = O(\varepsilon^{-2}p^{-2}) + O(\varepsilon^{-2}p^{-3+4\beta_{(1)}}). \tag{A. 24}$$

Let $\varepsilon = p^{\beta_{(1)}-1/2-\theta}$, $0 < \theta < 1/4$. By using Lemma 2, under (A1) and (A2),

$$\Pr\left\{H_{\ell m} \le f_{\ell m}(x) + (4/3)\eta(\varepsilon)(\lambda^2+1)\right\} = \Phi(x) + o(p^{\beta_{(1)}-1/2}).$$

By combining this result with (A. 22)–(A. 24), under (A1) and (A2),

$$\begin{aligned}\Pr\left\{H_{\ell m} \le \widehat{f_{\ell m}}(x)\right\} \le& \Pr\left\{H_{\ell m} \le f_{\ell m}(x) + (4/3)\eta(\varepsilon)(\lambda^2+1)\right\} \\ &+ \Pr\left(|\widehat{b}_{\ell m}/\sigma^3_{\ell m} - b_{\ell m}/\sigma^3_{\ell m}| \ge \varepsilon\right) + \Pr\left(|\widehat{\sigma}^2_{\ell m}/\sigma^2_{\ell m} - 1| \ge \varepsilon\right) \\ =& \Phi(x) + o(p^{\beta_{(1)}-1/2}). \end{aligned}$$

A lower bound may be obtained in the same way, and so the theorem is proved.

E. Proof of Theorem 2

From Lemma 1 and Theorem 1, for each ℓ and m, under (A1) and (A2),

$$\begin{aligned}&\Pr\left(H_{\ell m} \ge \widehat{\sigma}_{\ell m} z_{\alpha_{\mathrm{pw}}}\right) = \alpha_{\mathrm{pw}} + O(p^{\beta_{(1)}-1/2}),\ \Pr\left\{H_{\ell m} \ge \widehat{\sigma}_{\ell m}\widehat{f_{\ell m}}(z_{\alpha_{\mathrm{pw}}})\right\} = \alpha_{\mathrm{pw}} + o(p^{\beta_{(1)}-1/2}), \\ &\Pr\left(H_{1m} \ge \widehat{\sigma}_{1m} z_{\alpha_{\mathrm{c}}}\right) = \alpha_{\mathrm{c}} + O(p^{\beta_{(1)}-1/2}),\ \Pr\left\{H_{1m} \ge \widehat{\sigma}_{1m}\widehat{f_{1m}}(z_{\alpha_{\mathrm{c}}})\right\} = \alpha_{\mathrm{c}} + o(p^{\beta_{(1)}-1/2}).\end{aligned}$$

Substituting these asymptotic results into L_{pw1}, L_{pw2}, L_{c1}, and L_{c2} proves the theorem.

Acknowledgments

We would like to thank Editage (www.editage.com) for English language editing. Also, the first author is grateful to his late domestic cat, Uribo, for offering needed stimulus and energizing vitality during study times despite suffering from cancer.

The research of the first author was supported by JSPS KAKENHI Grant Number 20K11712. The research of the second author was supported by JSPS KAKENHI Grant Number 20K11714 and general support of Senshu Research Abroad Program (FY 2020).

References

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

[2] Bretz, F., Hothorn, T., Westfall, P., 2010. Multiple comparisons using R. Chapman & Hall/CRC.

[3] Chen, S.X., Qin, Y.L., 2010. A two-sample test for high-dimensional data with applications to gene-set testing. Ann Statist. 38, 808–835.

[4] Dempster, A.P., 1958. A high dimensional two samples significance test. Ann. Math. Stat. 29, 995–1010.

[5] Dempster, A.P., 1960. A significance test for separation of two highly multivariate small samples. Biometrics. 16, 41–50.

[6] Feng, L., Zou, C., Wang, Z., Zhu, L., 2015. Two sample Behrens-Fisher problem for high-dimensional data. Statist. Sinica 25, 1297-1312.

[7] Hall, P. 1983. Inverting an edgeworth expansion. Ann Statist. 11, 569–576.

[8] Hochberg, Y., Tamhane, A.C., 1987. Multiple comparison Procedures. Wiley, New York.

[9] Hsu, J.C., 1996. Multiple comparisons. Theory and methods. Chapman & Hall, London.

[10] Hu, J., Bai, Z., Wang, C., Wang, W., 2017. On testing the equality of high dimensional mean vectors with unequal covariance matrices. Ann. Inst. Statist. Math. 69, 365-387.

[11] Ishii, A., Yata, K., Aoshima, M., 2019. Inference on high-dimensional mean vectors under the strongly spiked eigenvalue model. Jpn. J. Stat. Data Sci. 2, 105-128.

[12] Hyodo, M., Takahashi, S., Nishiyama, T., 2014. Multiple comparisons among mean vectors when the dimension is large than the total sample size. Commun. Stat.-Simul. Comput. 43, 2283–2306.

[13] Mansouri, K., Ringsted, T., Ballabio, D., Todeschini, R., Consonni, V., 2013. Quantitative structure-activity relationship models for ready biodegradability of chemicals. J. Chem. Inf. Model. 53, 867–878.

[14] Nishiyama, T., Hyodo, M., Seo, T., Pavlenko, T., 2013. Testing linear hypotheses of mean vectors for high-dimension data with unequal covariance matrices. J. Stat. Plann. Inference. 143, 1898–1911.

[15] Seo, T., Nishiyama, T., 2008. On the conservative simultaneous confidence procedures for multiple comparisons among mean vectors. J. Stat. Plann. Inference. 138, 3448–3456.

[16] Seo, T., Mano, S., Fujikoshi, Y., 1994. A generalized Tukey conjecture for multiple comparisons among mean vectors. J. Amer. Statist. Assoc. 89, 676–679.

[17] Seo, T., Siotani, M., 1992. The multivariate studentized range and its upper percentiles. J. Japan Statist. Soc. 22, 123–137.

[18] Takahashi, S., Hyodo, M., Nishiyama, T., Pavlenko, T., 2013. Multiple comparison procedures for high-dimensional data and their robustness under non-normality. J. Japanese Soc. Comput. Statist. 26, 71–82.

[19] Zhang, J.-T., Zhou, B., Guo, J., Zhu, T., 2021. Two-sample Behrens-Fisher problems for high-dimensional data: A normal reference approach. J. Stat. Plann. Inference. 213, 142–161.