MULTIPLE COMPARISON PROCEDURES FOR HIGH-DIMENSIONAL DATA AND THEIR ROBUSTNESS UNDER NON-NORMALITY

Sho Takahashi*, Masashi Hyodo†, Takahiro Nishiyama‡ and Tatjana Pavlenko§

ABSTRACT

This paper analyzes whether procedures for multiple comparison derived in Hyodo *et al.* (2013) work for an unbalanced case and under non-normality. We focus on pairwise multiple comparisons and comparisons with a control among mean vectors, and show that the asymptotic properties of these procedures remain valid in an unbalanced high-dimensional setting. We also numerically justify that the derived procedures are robust under non-normality, i.e., the coverage probability of these procedures can be controlled with or without the assumption of normality of the data.

1. Introduction and motivation

In this paper, we consider pairwise multiple comparisons and comparisons with a control among mean vectors in the unbalanced case, i.e., assuming unequal sample sizes. Let $\mathbf{x}_j^{(i)}$ ($j = 1, 2, \dots, N_i$, $i = 1, 2, \dots, k$) be independently distributed as the p -dimensional normal distribution with mean vector $\boldsymbol{\mu}^{(i)}$ and common covariance matrix Σ . Let the i -th sample mean vector and the pooled sample covariance matrix be

$$\bar{\mathbf{x}}^{(i)} = \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{x}_j^{(i)}, \quad S = \frac{1}{m} \sum_{i=1}^k \sum_{j=1}^{N_i} (\mathbf{x}_j^{(i)} - \bar{\mathbf{x}}^{(i)})(\mathbf{x}_j^{(i)} - \bar{\mathbf{x}}^{(i)})',$$

respectively, where $m = \sum_{i=1}^k N_i - k$.

In general, the simultaneous confidence intervals for pairwise multiple comparisons among mean vectors with the confidence level $1 - \alpha$ are given by

$$\mathbf{a}'(\boldsymbol{\mu}^{(i)} - \boldsymbol{\mu}^{(j)}) \in \left[\mathbf{a}'(\bar{\mathbf{x}}^{(i)} - \bar{\mathbf{x}}^{(j)}) \pm t_p \sqrt{w_{ij} \mathbf{a}' S \mathbf{a}} \right], \quad \forall \mathbf{a} \in \mathbf{R}^p - \{\mathbf{0}\}, \quad 1 \leq i < j \leq k,$$

where $w_{ij} = 1/N_i + 1/N_j$, $\mathbf{R}^p - \{\mathbf{0}\}$ is the set of any nonnull real p -dimensional vectors, and $t_p^2 \equiv t_p^2(\alpha)$ is the upper 100α percentile of the distribution of $T_{\max \cdot p}^2$ defined by

$$\Pr\{T_{\max \cdot p}^2 > t_p^2\} = \alpha,$$

*Chiba University Hospital Clinical Research Center, 1-8-1, Inohana, Chuo-ku, Chiba, 260-8677, Japan
E-mail: sho@chiba-u.jp

†Department of Mathematical Information Science, Tokyo University of Science, 1-3, Kagurazaka, Shinjuku-ku, Tokyo, 162-8601, Japan

‡Department of Business Administration, Senshu University, 2-1-1, Higashimita, Tama-ku, Kawasaki-shi, Kanagawa, 214-8580, Japan

§Department of Mathematics, KTH Royal Institute of Technology, 106-91, Stockholm, Sweden

Key words: Asymptotic expansion; Bonferroni's inequality; Comparison with a control; Dempster trace criterion; High-dimensional data; Monte Carlo Simulation; Pairwise comparison; Power, Robustness

where

$$\begin{aligned} T_{\max \cdot p}^2 &= \max_{1 \leq i < j \leq k} \{T_{ij}^2\}, \\ T_{ij}^2 &= w_{ij}^{-1}(\mathbf{y}^{(i)} - \mathbf{y}^{(j)})' S^{-1}(\mathbf{y}^{(i)} - \mathbf{y}^{(j)}), \\ \mathbf{y}^{(i)} &= \bar{\mathbf{x}}^{(i)} - \boldsymbol{\mu}^{(i)}, \quad i = 1, 2, \dots, k. \end{aligned}$$

By analogy with the above, in the case of comparisons with a control, letting the first population be a control, the simultaneous confidence intervals are given by

$$\mathbf{a}'(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(j)}) \in \left[\mathbf{a}'(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(j)}) \pm t_c \sqrt{w_{1j} \mathbf{a}' S \mathbf{a}} \right], \quad \forall \mathbf{a} \in \mathbf{R}^p - \{\mathbf{0}\}, \quad 2 \leq j \leq k,$$

and the value $t_c^2 \equiv t_c^2(\alpha)$ is chosen to satisfy

$$\Pr\{T_{\max \cdot c}^2 > t_c^2\} = \alpha,$$

where

$$\begin{aligned} T_{\max \cdot c}^2 &= \max_{2 \leq j \leq k} \{T_{1j}^2\}, \\ T_{1j}^2 &= w_{1j}^{-1}(\mathbf{y}^{(1)} - \mathbf{y}^{(j)})' S^{-1}(\mathbf{y}^{(1)} - \mathbf{y}^{(j)}). \end{aligned}$$

In order to construct these simultaneous confidence intervals, it is required to obtain the upper percentiles of $T_{\max}^2$ -type statistics, i.e., t_p and t_c . In general, it is difficult to obtain them exactly even under normality. The use of the Bonferroni's inequality to obtain approximate upper percentiles is discussed by e.g., Seo and Siotani (1992), Seo (1995). These results are derived for the standard asymptotic setting, i.e., assuming that p is fixed and is much smaller than N_i .

However, it is well known that when the dimension is larger than the total sample size, the sample covariance matrix becomes singular, and hence it will be impossible to define Hotelling's $T_{\max}^2$ -type statistic. To tackle this problem efficiently, the Dempster trace criterion for one and two samples can be used.

The technique considered in the current study develops results derived in Dempster (1958, 1960). A similar approach for multivariate linear hypotheses has been also discussed by Fujikoshi *et al.* (2004), Himeno (2007), Nishiyama *et al.* (2013) and many other authors.

To adjust the high-dimensional setting to the unbalanced case, we consider the following asymptotic framework:

$$\begin{aligned} \text{A1} : m &\rightarrow \infty, \quad p \rightarrow \infty, \quad p/m \rightarrow \gamma \in (0, \infty), \\ \text{A2} : 0 &< \lim_{p \rightarrow \infty} \text{tr} \Sigma^i / p < \infty, \quad i = 1, 2, \dots, 8. \end{aligned}$$

Using conventional terminology, from now on we will refer to the assumption A1 as (m, p) -asymptotics.

In Hyodo *et al.* (2013), the balanced case is considered and the following $D_{\max}$ -type test statistic

$$\begin{aligned} D_{\max \cdot p} &= \max_{1 \leq i < j \leq k} \{D_{ij}\}, \\ D_{ij} &= \frac{p}{\hat{\sigma}} \left\{ \frac{W_{ij}}{\text{tr} S} - 1 \right\} = \frac{p}{\hat{\sigma}} \left\{ \frac{(\mathbf{y}^{(i)} - \mathbf{y}^{(j)})' (\mathbf{y}^{(i)} - \mathbf{y}^{(j)})}{w_{ij} \text{tr} S} - 1 \right\} \end{aligned} \quad (1)$$

is proposed for pairwise comparisons. Further, for comparisons with a control, the following statistic

$$D_{\max \cdot c} = \max_{2 \leq j \leq k} \{D_{1j}\},$$

$$D_{1j} = \frac{p}{\hat{\sigma}} \left\{ \frac{W_{1j}}{\text{tr} S} - 1 \right\} = \frac{p}{\hat{\sigma}} \left\{ \frac{(\mathbf{y}^{(1)} - \mathbf{y}^{(j)})'(\mathbf{y}^{(1)} - \mathbf{y}^{(j)})}{w_{1j} \text{tr} S} - 1 \right\} \quad (2)$$

is derived, where $\hat{\sigma} = \sqrt{2p\hat{c}_2/\hat{c}_1^2}$ and $\hat{c}_i$'s are the unbiased and consistent estimators of $c_i = \text{tr} \Sigma^i / p$. Further, the Bonferroni approximations for the upper percentiles of these statistics are derived.

This study gives an extension of the results by Hyodo *et al.* (2013) to the unbalanced case. We show that the consistency and asymptotic normality of D_{ij} in (1) and D_{1j} in (2) remain valid for the unbalanced case. For the procedures derived in Hyodo *et al.* (2013), the coverage probability can be controlled provided the assumption that the data are normally distributed. However, so far no satisfactory test is available to ascertain the multivariate normality of the data when $N_i \leq p$. Hence, it would be desirable to derive such multiple comparison procedures for which the coverage probability can be controlled with or without the assumption of normality of the data, that is, their procedures are robust under non-normality. Unfortunately, it would be difficult to derive such procedures theoretically. Our objective in this study is to numerically justify that the extended comparison procedures are robust for the model described above.

The rest of the paper is organized as follows: In Section 2, the main asymptotic properties are shown to be valid for the unbalanced case. Further, the Bonferroni approximations of the upper 100α percentiles of $D_{\max}$ -type statistics are derived for pairwise comparisons and comparisons with a control in Section 2. Section 3 provides numerical examination of the performance accuracy of the extended procedures. To justify the robustness of these procedures, we compare the power of $D_{\max}$ -type statistics with that of the data generated from a number of non-normal distributions. Finally, we provide some concluding remarks.

2. Multiple comparisons in high-dimensional framework – unbalanced case

2.1. Pairwise comparisons

Consider the following simultaneous confidence intervals for mean vectors:

$$\mathbf{a}'(\boldsymbol{\mu}^{(i)} - \boldsymbol{\mu}^{(j)}) \in \left[\mathbf{a}'(\bar{\mathbf{x}}^{(i)} - \bar{\mathbf{x}}^{(j)}) \pm d_p \sqrt{w_{ij}(\text{tr} S) \mathbf{a}' \mathbf{a}} \right], \quad \forall \mathbf{a} \in \mathbf{R}^p - \{\mathbf{0}\}, \quad 1 \leq i < j \leq k,$$

where $d_p^2 = 1 + (\hat{\sigma}/p)z_p$ and $z_p \equiv z_p(\alpha)$ is the upper 100α percentile of the $D_{\max \cdot p}$ statistic. In a high-dimensional framework, it is difficult to give the exact value of z_p . In this study, we derive the approximation for z_p using the Bonferroni's inequality.

By applying Bonferroni's inequality to $\Pr\{D_{\max \cdot p} > z_p\}$, we get

$$\Pr\{D_{\max \cdot p} > z_p\} < \sum_{i=1}^{k-1} \sum_{j=i+1}^k \Pr\{D_{ij} > z_p\}.$$

We then define the Bonferroni approximation for z_p as such $z_{1,p}$ which satisfies

$$\sum_{i=1}^{k-1} \sum_{j=i+1}^k \Pr\{D_{ij} > z_{1,p}\} = \alpha$$

for a given α . It is well known that $z_{1,p}$ essentially overestimates the upper 100α percentile of $D_{\max \cdot p}$. In order to get a corrected estimator of this percentile, we derive the Cornish-Fisher expansion of $z_{1,p}$. This expansion is derived in Hyodo *et al.* (2013) in the balanced case. In a similar manner, we can show that the result of Hyodo *et al.* (2013) is applicable to the unbalanced case. Thus, under the high-dimensional asymptotic frameworks A1 and A2, we can obtain the following approximation for $z_{1,p}$ in the unbalanced case:

$$\hat{z}_{1,p} = z_{\alpha_p} + \frac{1}{\sqrt{p}} \left(\frac{\sqrt{2}\hat{c}_3}{3\sqrt{\hat{c}_2^3}} \right) (z_{\alpha_p}^2 - 1) + \frac{1}{p} \left\{ \frac{\hat{c}_4}{2\hat{c}_2^2} z_{\alpha_p} (z_{\alpha_p}^2 - 3) - \frac{2\hat{c}_3^2}{9\hat{c}_2^3} z_{\alpha_p} (2z_{\alpha_p}^2 - 5) \right\} + \frac{1}{2m} z_{\alpha_p},$$

where z_α is the upper 100α percentile of the standard normal distribution, $\alpha_p = \alpha/K$ and $K = k(k-1)/2$. Also $\hat{c}_i$'s ($i = 1, 2, 3, 4$) are the unbiased and (m, p) -consistent estimators of c_i with $\text{Var}[\hat{c}_i] = O(m^{-2})$ as follows:

$$\begin{aligned} \hat{c}_1 &= \frac{\text{tr}S}{p}, \\ \hat{c}_2 &= \frac{m^2}{(m+2)(m-1)p} \left\{ \text{tr}S^2 - \frac{(\text{tr}S)^2}{m} \right\}, \\ \hat{c}_3 &= \frac{m^4}{(m+4)(m+2)(m-1)(m-2)p} \left\{ \text{tr}S^3 - \frac{3}{m} \text{tr}S^2 \text{tr}S + \frac{2}{m^2} (\text{tr}S)^3 \right\}, \\ \hat{c}_4 &= \frac{m^3}{(m+6)(m+4)(m+2)(m+1)(m-1)(m-2)(m-3)p} \\ &\quad \times \left[b_1 \text{tr}S^4 + b_2 \text{tr}S^3 \text{tr}S + b_3 (\text{tr}S^2)^2 + b_4 \text{tr}S^2 (\text{tr}S)^2 + b_5 (\text{tr}S)^4 \right] \end{aligned}$$

where

$$\begin{aligned} b_1 &= m^2(m^2 + m + 2), \quad b_2 = -4m(m^2 + m + 2), \quad b_3 = -m(2m^2 + 3m - 6), \\ b_4 &= 2m(5m + 6), \quad b_5 = -(5m + 6). \end{aligned}$$

2.2. Comparisons with a control

The simultaneous confidence intervals for comparisons with a control are given by

$$\mathbf{a}'(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(j)}) \in \left[\mathbf{a}'(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(j)}) \pm d_c \sqrt{w_{1j}(\text{tr}S) \mathbf{a}' \mathbf{a}} \right], \quad \forall \mathbf{a} \in \mathbf{R}^p - \{\mathbf{0}\}, \quad 2 \leq j \leq k,$$

where $d_c^2 = 1 + (\hat{\sigma}/p)z_c$ and $z_c \equiv z_c(\alpha)$ is the upper 100α percentile of the $D_{\max \cdot c}$ statistic.

In a similar manner to $D_{\max \cdot p}$, under the high-dimensional and asymptotic frameworks A1 and A2, the Bonferroni approximation for z_c can be obtained as

$$\hat{z}_{1,c} = z_{\alpha_c} + \frac{1}{\sqrt{p}} \left(\frac{\sqrt{2}\hat{c}_3}{3\sqrt{\hat{c}_2^3}} \right) (z_{\alpha_c}^2 - 1) + \frac{1}{p} \left\{ \frac{\hat{c}_4}{2\hat{c}_2^2} z_{\alpha_c} (z_{\alpha_c}^2 - 3) - \frac{2\hat{c}_3^2}{9\hat{c}_2^3} z_{\alpha_c} (2z_{\alpha_c}^2 - 5) \right\} + \frac{1}{2m} z_{\alpha_c},$$

where $\alpha_c = \alpha/(k-1)$.

3. Numerical examinations

We conduct Monte Carlo simulation with $\ell = 100000$ replications, and numerically evaluate the accuracy of approximations and the robustness under non-normality in terms of the coverage probability.

3.1. Accuracy of approximations

In this subsection, we evaluate the coverage probability of the proposed procedures. The selected parameter values are $\alpha = 0.05$ and $p = 30, 60, 90, 120, 150, 200$. The sample sizes are set as $(N_1, N_2, N_3) = (20, 20, 20), (30, 20, 10), (40, 10, 10)$ when $k = 3$. Also let $N_{12} \equiv N_1 = N_2$, $N_{34} \equiv N_3 = N_4$ and $N_{56} \equiv N_5 = N_6$ when $k = 6$, then the sample sizes are set as $(N_{12}, N_{34}, N_{56}) = (10, 10, 10), (15, 10, 5), (20, 5, 5)$. The covariance structures are as follows: $\Sigma_1 = I_p$, $\Sigma_2 = (0.2^{|i-j|})$ and $\Sigma_3 = (0.5^{|i-j|})$. Attained confidence levels (i.e., empirical coverage probability) of $D_{\max}$ -type statistics can be numerically evaluated as

$$Q_p(x) = \frac{\#\{D_{\max \cdot p} < x\}}{\ell},$$

$$Q_c(x) = \frac{\#\{D_{\max \cdot c} < x\}}{\ell}.$$

Tables 1 and 2 list the simulated value z_p^* (i.e., $Q_p(z_p^*) = 1 - \alpha$) and the attained confidence level $Q_p(x)$ for pairwise comparisons. It should be noted that $Q_p(\hat{z}_{1 \cdot p}) \geq 1 - \alpha$. $\hat{z}_{1 \cdot p}$ has a tendency to be more conservative as the number of populations increases. z_p^* is larger for the balanced case and $Q_p(\hat{z}_{1 \cdot p})$ turns out to be more conservative for the unbalanced case.

Tables 3 and 4 list the value z_c^* (i.e., $Q_c(z_c^*) = 1 - \alpha$) and the attained confidence level $Q_c(x)$ for comparisons with a control. In contrast with pairwise comparisons, the balanced case is more conservative than the unbalanced cases.

Table 1: The attained confidence level for pairwise comparisons when $k = 3$.

				Σ_1			Σ_2			Σ_3		
N_1	N_2	N_3	p	z_p^*	$Q_p(z_{\alpha_p})$	$Q_p(\hat{z}_{1 \cdot p})$	z_p^*	$Q_p(z_{\alpha_p})$	$Q_p(\hat{z}_{1 \cdot p})$	z_p^*	$Q_p(z_{\alpha_p})$	$Q_p(\hat{z}_{1 \cdot p})$
20	20	20	30	2.414	0.920	0.952	2.440	0.918	0.952	2.599	0.905	0.952
			60	2.342	0.928	0.951	2.362	0.925	0.951	2.474	0.916	0.952
			90	2.295	0.932	0.952	2.320	0.929	0.952	2.400	0.922	0.953
			120	2.291	0.932	0.950	2.285	0.933	0.953	2.370	0.924	0.953
			150	2.259	0.936	0.952	2.264	0.935	0.953	2.353	0.926	0.952
			200	2.237	0.938	0.952	2.242	0.937	0.953	2.321	0.929	0.952
30	20	10	30	2.398	0.923	0.953	2.443	0.919	0.952	2.597	0.907	0.952
			60	2.319	0.930	0.953	2.339	0.928	0.954	2.456	0.918	0.953
			90	2.288	0.933	0.953	2.298	0.932	0.953	2.403	0.922	0.953
			120	2.259	0.936	0.953	2.290	0.932	0.952	2.354	0.926	0.954
			150	2.230	0.938	0.955	2.256	0.936	0.954	2.340	0.928	0.953
			200	2.233	0.939	0.953	2.234	0.938	0.954	2.311	0.930	0.953
40	10	10	30	2.393	0.923	0.954	2.441	0.918	0.952	2.599	0.908	0.952
			60	2.315	0.929	0.953	2.343	0.928	0.953	2.450	0.920	0.953
			90	2.278	0.934	0.954	2.284	0.932	0.954	2.399	0.923	0.953
			120	2.255	0.936	0.954	2.278	0.934	0.953	2.352	0.927	0.954
			150	2.253	0.937	0.953	2.256	0.936	0.954	2.325	0.929	0.954
			200	2.229	0.939	0.953	2.239	0.938	0.953	2.298	0.932	0.954

Table 2: The attained confidence level for pairwise comparisons when $k = 6$.

				Σ_1			Σ_2			Σ_3		
N_{12}	N_{34}	N_{56}	p	z_p^*	$Q_p(z_{\alpha_p})$	$Q_p(\widehat{z}_{1 \cdot p})$	z_p^*	$Q_p(z_{\alpha_p})$	$Q_p(\widehat{z}_{1 \cdot p})$	z_p^*	$Q_p(z_{\alpha_p})$	$Q_p(\widehat{z}_{1 \cdot p})$
10	10	10	30	3.250	0.876	0.952	3.318	0.868	0.952	3.635	0.834	0.953
			60	3.093	0.897	0.953	3.145	0.890	0.952	3.384	0.861	0.951
			90	3.013	0.907	0.954	3.062	0.901	0.953	3.247	0.878	0.953
			120	2.961	0.915	0.955	3.013	0.908	0.953	3.177	0.885	0.953
			150	2.945	0.918	0.954	2.976	0.912	0.954	3.134	0.892	0.952
			200	2.909	0.922	0.954	2.950	0.917	0.952	3.062	0.901	0.954
15	10	5	30	3.224	0.882	0.955	3.303	0.871	0.954	3.609	0.838	0.955
			60	3.071	0.900	0.955	3.112	0.896	0.955	3.351	0.867	0.954
			90	2.997	0.911	0.956	3.038	0.905	0.955	3.227	0.881	0.955
			120	2.961	0.916	0.955	2.989	0.913	0.955	3.161	0.890	0.954
			150	2.935	0.919	0.955	2.953	0.916	0.956	3.111	0.897	0.954
			200	2.902	0.924	0.955	2.911	0.921	0.956	3.058	0.903	0.954
20	5	5	30	3.207	0.884	0.956	3.290	0.876	0.955	3.609	0.842	0.955
			60	3.053	0.905	0.956	3.110	0.896	0.956	3.337	0.871	0.956
			90	2.990	0.913	0.956	3.033	0.907	0.955	3.213	0.885	0.956
			120	2.945	0.918	0.956	2.997	0.913	0.955	3.150	0.893	0.955
			150	2.923	0.922	0.956	2.946	0.918	0.956	3.101	0.899	0.956
			200	2.891	0.927	0.956	2.905	0.925	0.957	3.044	0.906	0.956

Table 3: The attained confidence level for comparisons with a control when $k = 3$.

				Σ_1			Σ_2			Σ_3		
N_1	N_2	N_3	p	z_c^*	$Q_c(z_{\alpha_c})$	$Q_c(\widehat{z}_{1 \cdot c})$	z_c^*	$Q_c(z_{\alpha_c})$	$Q_c(\widehat{z}_{1 \cdot c})$	z_c^*	$Q_c(z_{\alpha_c})$	$Q_c(\widehat{z}_{1 \cdot c})$
20	20	20	30	2.197	0.927	0.951	2.221	0.926	0.951	2.330	0.918	0.952
			60	2.123	0.933	0.952	2.152	0.931	0.951	2.241	0.924	0.951
			90	2.095	0.936	0.952	2.133	0.933	0.950	2.203	0.927	0.950
			120	2.094	0.936	0.950	2.091	0.936	0.952	2.148	0.931	0.952
			150	2.062	0.939	0.952	2.076	0.938	0.952	2.131	0.933	0.952
			200	2.056	0.940	0.951	2.065	0.938	0.952	2.112	0.935	0.952
30	20	10	30	2.223	0.924	0.949	2.243	0.923	0.949	2.367	0.914	0.949
			60	2.138	0.931	0.950	2.168	0.928	0.949	2.279	0.920	0.948
			90	2.110	0.934	0.950	2.132	0.932	0.950	2.202	0.926	0.950
			120	2.094	0.936	0.950	2.115	0.934	0.949	2.176	0.927	0.950
			150	2.079	0.937	0.950	2.096	0.935	0.950	2.166	0.929	0.949
			200	2.063	0.938	0.951	2.066	0.938	0.951	2.136	0.932	0.950
40	10	10	30	2.236	0.922	0.947	2.254	0.921	0.948	2.379	0.913	0.948
			60	2.146	0.931	0.950	2.149	0.930	0.951	2.253	0.921	0.950
			90	2.127	0.932	0.949	2.136	0.931	0.949	2.218	0.924	0.949
			120	2.090	0.936	0.950	2.119	0.934	0.949	2.183	0.927	0.949
			150	2.078	0.937	0.951	2.102	0.934	0.950	2.160	0.929	0.949
			200	2.070	0.938	0.950	2.082	0.937	0.950	2.122	0.933	0.951

Further, we conduct a Monte Carlo simulation for the approximate upper percentiles of the distribution of $D_{12} = D_{ij}$ in terms of the attained significance level (i.e., type I error rate). This simulation is performed to examine the conservativeness of our procedures by dividing into the effects of the approximation error by the Cornish-Fisher expansion and that

Multiple Comparisons for High-Dimensional Data

Table 4: The attained confidence level for comparisons with a control when $k = 6$.

				Σ_1			Σ_2			Σ_3		
N_{12}	N_{34}	N_{56}	p	z_c^*	$Q_c(z_{\alpha_c})$	$Q_c(\widehat{z}_{1 \cdot c})$	z_c^*	$Q_c(z_{\alpha_c})$	$Q_c(\widehat{z}_{1 \cdot c})$	z_c^*	$Q_c(z_{\alpha_c})$	$Q_c(\widehat{z}_{1 \cdot c})$
10	10	10	30	2.716	0.904	0.950	2.730	0.905	0.952	2.934	0.888	0.953
			60	2.565	0.922	0.954	2.620	0.918	0.952	2.768	0.902	0.952
			90	2.512	0.928	0.955	2.562	0.923	0.952	2.674	0.912	0.953
			120	2.499	0.929	0.953	2.524	0.927	0.953	2.636	0.916	0.953
			150	2.493	0.930	0.952	2.493	0.930	0.954	2.610	0.918	0.952
			200	2.452	0.935	0.954	2.468	0.933	0.954	2.555	0.923	0.954
15	10	5	30	2.716	0.904	0.950	2.759	0.899	0.950	2.957	0.882	0.951
			60	2.589	0.917	0.952	2.624	0.915	0.951	2.795	0.896	0.950
			90	2.554	0.922	0.951	2.584	0.919	0.951	2.712	0.905	0.950
			120	2.519	0.925	0.952	2.550	0.922	0.951	2.650	0.910	0.951
			150	2.503	0.928	0.951	2.517	0.925	0.952	2.622	0.914	0.951
			200	2.487	0.930	0.951	2.497	0.929	0.951	2.581	0.919	0.951
20	5	5	30	2.723	0.903	0.949	2.775	0.897	0.949	2.986	0.875	0.949
			60	2.613	0.915	0.950	2.638	0.910	0.950	2.808	0.892	0.949
			90	2.574	0.919	0.949	2.579	0.919	0.951	2.724	0.903	0.949
			120	2.529	0.924	0.950	2.537	0.923	0.952	2.669	0.908	0.950
			150	2.509	0.927	0.951	2.531	0.924	0.950	2.646	0.911	0.949
			200	2.484	0.930	0.951	2.506	0.928	0.950	2.591	0.917	0.950

of the overestimate by the Bonferroni's inequality. The selected parameter values are $\alpha = 0.05$, $p = 30, 60, 90, 120, 150, 200$, $(N_1, N_2) = (20, 20), (30, 10)$, and $\Sigma_1 = I_p$, $\Sigma_2 = (0.2^{|i-j|})$ and $\Sigma_3 = (0.5^{|i-j|})$. Table 5 lists the results of the attained significance level of D_{12} based on z_α (Limit) and Cornish-Fisher (C-F) expansion. It can be observed from the results that the attained significance level of C-F is close to 0.05 and the approximation is good. Therefore, the conservativeness of our procedures may be caused by the Bonferroni's inequality. Also, in comparison the attained significance level for the balanced and unbalanced cases, the balanced case has a tendency to be slightly closer to 0.05 than the unbalanced case. From the above, the heterogeneity of the sample sizes has an effect on the approximation by the Cornish-Fisher expansion. Consequently, we suspect that the difference due to heterogeneity

Table 5: The attained significance level of D_{ij} .

			Σ_1		Σ_2		Σ_3	
N_1	N_2	p	Limit	C-F	Limit	C-F	Limit	C-F
20	20	30	0.065	0.051	0.067	0.051	0.071	0.053
		60	0.062	0.052	0.064	0.053	0.066	0.051
		90	0.061	0.051	0.060	0.051	0.065	0.052
		120	0.060	0.051	0.063	0.052	0.063	0.050
		150	0.059	0.050	0.060	0.050	0.062	0.051
		200	0.057	0.050	0.059	0.052	0.061	0.049
30	10	30	0.066	0.051	0.065	0.052	0.071	0.053
		60	0.062	0.052	0.065	0.053	0.067	0.051
		90	0.060	0.050	0.062	0.052	0.066	0.053
		120	0.060	0.051	0.060	0.052	0.064	0.052
		150	0.059	0.051	0.060	0.051	0.062	0.051
		200	0.059	0.050	0.059	0.052	0.062	0.052

of the sample size has occurred in coverage probability.

3.2. Robustness of the proposed procedures under non-normality

In this subsection, we evaluate the robustness of the proposed procedure under non-normality in terms of the coverage probability. In this context, we also evaluate the accuracy of the estimator $\hat{z}_{1,p}$ and the approximation error by the Cornish-Fisher expansion under non-normality.

To investigate how sensitive our procedure is to the normality assumption, we study the following five distributions:

D1: the multivariate normal distribution,

D2: the multivariate t distribution with 7 degrees of freedom,

D3: the ε -contaminated normal distribution ($\kappa = 1.78$),

D4: the ε -contaminated normal distribution ($\kappa = 3.24$),

D5: the multivariate skew-normal distribution ($\delta = \delta_j = 0.5$, $j = 1, 2, \dots, p$).

It should be noted that D1–D4 belong to the class of elliptical distributions and symmetric distributions, whereas D5 represents the case of an asymmetric distribution. Parameters are the same settings in Subsection 3.1.

Table 6 lists the attained confidence levels for pairwise comparisons. When the sample sizes are balanced, all values are greater than or equal to 0.95. The values in D2–D4 are large compared to that in D1. When $(N_1, N_2, N_3) = (30, 20, 10)$, the same tendency as the balanced case can be observed. Also when $(N_1, N_2, N_3) = (40, 10, 10)$, D1, D3 and D5 are greater than or equal to 0.95, however, D2 and D3 are less than 0.95. Further, D1 and D5 are just slightly affected by whether the sample sizes are balanced or unbalanced, in

Table 6: The robustness for pairwise comparisons.

N_1	N_2	N_3	p	D1	D2	D3	D4	D5
20	20	20	30	0.952	0.972	0.958	0.965	0.953
			60	0.951	0.982	0.960	0.965	0.952
			90	0.952	0.987	0.959	0.964	0.953
			120	0.950	0.991	0.959	0.965	0.952
			150	0.952	0.993	0.960	0.964	0.951
			200	0.952	0.996	0.961	0.965	0.952
30	20	10	30	0.953	0.959	0.955	0.956	0.953
			60	0.954	0.963	0.956	0.957	0.953
			90	0.953	0.965	0.955	0.957	0.952
			120	0.953	0.968	0.955	0.956	0.953
			150	0.954	0.969	0.955	0.957	0.954
			200	0.954	0.970	0.956	0.957	0.953
40	10	10	30	0.954	0.947	0.950	0.948	0.953
			60	0.953	0.944	0.950	0.949	0.954
			90	0.954	0.943	0.951	0.948	0.955
			120	0.954	0.942	0.951	0.949	0.953
			150	0.953	0.940	0.951	0.947	0.954
			200	0.953	0.940	0.950	0.948	0.954

contrast to D2–D4 which are strongly influenced by the sample sizes. In particular, D2 can be expected to be very sensitive to high-dimensions due to the heavy tail of the multivariate t distribution with low degrees of freedom.

Table 7 lists the attained confidence levels for comparisons with a control. Almost all the values are greater than or equal to 0.95 when $(N_1, N_2, N_3) = (20, 20, 20), (30, 20, 10)$. In contrast, almost all the values are less than or equal to 0.95 when $(N_1, N_2, N_3) = (40, 10, 10)$.

In analogy with Subsection 3.1, we also examine our procedures by dividing the approximation error into parts by the Cornish-Fisher expansion and the overestimate by the Bonferroni's inequality under non-normality (D2–D5). The selected parameter values are $\alpha = 0.05$, $p = 30, 60, 90, 120, 150, 200$, $(N_1, N_2) = (20, 20), (30, 10)$. To evaluate them, we conduct a Monte Carlo simulation with 10000 replications. Table 8 lists the results in D1–

Table 7: The robustness for comparisons with a control.

N_1	N_2	N_3	p	D1	D2	D3	D4	D5
20	20	20	30	0.951	0.969	0.958	0.961	0.949
			60	0.952	0.978	0.958	0.963	0.950
			90	0.952	0.984	0.957	0.962	0.951
			120	0.950	0.988	0.958	0.962	0.953
			150	0.952	0.991	0.957	0.962	0.951
			200	0.951	0.994	0.958	0.962	0.951
30	20	10	30	0.949	0.956	0.952	0.953	0.951
			60	0.950	0.958	0.953	0.955	0.950
			90	0.950	0.960	0.953	0.955	0.950
			120	0.950	0.962	0.951	0.953	0.950
			150	0.950	0.963	0.953	0.953	0.951
			200	0.951	0.964	0.953	0.955	0.951
40	10	10	30	0.947	0.938	0.945	0.941	0.950
			60	0.950	0.931	0.945	0.940	0.949
			90	0.949	0.927	0.944	0.941	0.951
			120	0.950	0.925	0.945	0.941	0.950
			150	0.951	0.923	0.946	0.943	0.949
			200	0.950	0.919	0.945	0.942	0.950

Table 8: The attained significance level of D_{ij} under non-normality.

N_1	N_2	p	D1	D2	D3	D4	D5
20	20	30	0.055	0.024	0.043	0.038	0.051
		60	0.051	0.017	0.043	0.037	0.052
		90	0.048	0.007	0.039	0.041	0.053
		120	0.051	0.008	0.041	0.033	0.053
		150	0.051	0.005	0.042	0.035	0.050
		200	0.051	0.003	0.044	0.034	0.050
30	10	30	0.050	0.048	0.048	0.048	0.053
		60	0.052	0.042	0.042	0.049	0.055
		90	0.052	0.039	0.053	0.049	0.051
		120	0.051	0.038	0.048	0.046	0.052
		150	0.052	0.036	0.049	0.046	0.048
		200	0.049	0.037	0.047	0.044	0.051

D5. The attained significance levels exhibit the same tendency in D1. Also D2–D4 are always less than 0.05. In particular, D2 and D4 may be sensitive to heterogeneity of the sample size. Consequently, we suspect that the difference due to heterogeneity of the sample size has occurred in coverage probability even under non-normality.

Further, we evaluate the accuracy of the estimator $\hat{z}_{1,p}$ in terms of the variance and the bias under non-normality. The selected parameters are $\alpha = 0.05$, $p = 30, 60, 90, 120, 150, 200$ and $M = N_1 + N_2 + N_3 = 60$ ($k = 3$). Table 9 lists the variances and the biases of $\hat{z}_{1,p}$ computed by Monte Carlo simulation with 10000 replications. In D1 and D5, the variance and the bias are very small and the variance tends to 0 as p increases. In D3 and D4, the variance is small, however, there exists a bias of the estimator $\hat{z}_{1,p}$. Also the bias becomes smaller as p increases. In contrast, the variance and the bias in D2 are larger than the others. Furthermore, they have a tendency to become large as p increases. Therefore, the results in D2 indicate that both the variance and the bias have an effect on the accuracy of estimator $\hat{z}_{1,p}$.

From these simulations for testing the robustness under non-normality, we see that the variability of the power results and the attained confidence level is most pronounced for D2. Therefore, our procedures could be sensitive to large p when the distribution underlying the data has heavy tails.

Table 9: The variance and the bias of the estimator $\hat{z}_{1,p}$ under non-normality.

Variance						
M	p	D1	D2	D3	D4	D5
60	30	4.175×10^{-4}	1.572×10^{-2}	1.032×10^{-3}	1.827×10^{-3}	4.126×10^{-4}
	60	2.646×10^{-4}	2.399×10^{-2}	4.825×10^{-4}	8.038×10^{-4}	2.627×10^{-4}
	90	2.102×10^{-4}	3.047×10^{-2}	3.517×10^{-4}	5.347×10^{-4}	2.093×10^{-4}
	120	1.818×10^{-4}	3.596×10^{-2}	2.797×10^{-4}	4.334×10^{-4}	1.816×10^{-4}
	150	1.623×10^{-4}	3.882×10^{-2}	2.332×10^{-4}	3.434×10^{-4}	1.603×10^{-4}
	200	1.430×10^{-4}	4.226×10^{-2}	2.014×10^{-4}	2.807×10^{-4}	1.413×10^{-4}
Bias						
M	p	D1	D2	D3	D4	D5
60	30	-0.002	0.146	0.037	0.064	0.001
	60	-0.001	0.206	0.027	0.047	0.000
	90	-0.001	0.246	0.023	0.039	0.000
	120	-0.001	0.280	0.020	0.034	0.000
	150	0.000	0.303	0.017	0.031	0.000
	200	-0.001	0.329	0.015	0.027	0.000

3.3. Power of the test

In this subsection, we numerically evaluate the power of the test. We re-express simultaneous confidence intervals as the following hypothesis testing problem:

$$H_{ij} : \boldsymbol{\mu}^{(i)} = \boldsymbol{\mu}^{(j)} \quad \text{vs.} \quad A_{ij} : \boldsymbol{\mu}^{(i)} \neq \boldsymbol{\mu}^{(j)}, \quad i < j.$$

Then we study the performance property of $D_{\max}$ -type statistics using a series of power simulations. Let $\alpha = 0.05$, $k = 3$, and mean vectors and covariance matrix be as follows: $\boldsymbol{\mu}^{(1)} = (\theta \mathbf{1}'_{p_1}, \mathbf{0}')'$ and $\boldsymbol{\mu}^{(2)} = \boldsymbol{\mu}^{(3)} = \mathbf{0}$, $\theta = 0.1, 0.3, 0.5, 0.7, 0.9$, $r = p_1/p = 0.25, 0.50, 0.75, 1.00$, and $\Sigma = I_p$. We set $p = 60$ and vary sample sizes as above and calculate the empirical power as

Multiple Comparisons for High-Dimensional Data

$$\hat{\beta}_p = \frac{\sharp\{D_{12} > \hat{z}_{1\cdot p}, D_{13} > \hat{z}_{1\cdot p}, D_{23} < \hat{z}_{1\cdot p} \text{ under } A_{12}, A_{13}, H_{23}\}}{\ell},$$

$$\hat{\beta}_c = \frac{\sharp\{D_{12} > \hat{z}_{1\cdot c}, D_{13} > \hat{z}_{1\cdot c} \text{ under } A_{12}, A_{13}\}}{\ell},$$

with $\hat{\beta}_p$ and $\hat{\beta}_c$ denoting the power of pairwise comparisons and that of comparisons with a control, respectively.

Table 10 lists the power for pairwise comparisons. The power of the test becomes greater as both r and θ increase. Further, it can be observed that the balanced case is the most powerful for all choices of θ and r .

Table 11 lists the powers for comparisons with a control, showing the same tendency as in the pairwise case.

Table 10: The power for pairwise comparisons.

N_1	N_2	N_3	θ	$r = 0.25$	$r = 0.50$	$r = 0.75$	$r = 1.00$
20	20	20	0.1	0.002	0.004	0.006	0.010
			0.3	0.056	0.310	0.669	0.887
			0.5	0.594	0.975	0.983	0.982
			0.7	0.973	0.983	0.983	0.982
			0.9	0.982	0.983	0.983	0.983
30	20	10	0.1	0.001	0.002	0.003	0.006
			0.3	0.036	0.235	0.539	0.777
			0.5	0.472	0.940	0.982	0.983
			0.7	0.936	0.983	0.982	0.982
			0.9	0.983	0.983	0.983	0.982
40	10	10	0.1	0.001	0.001	0.002	0.003
			0.3	0.018	0.144	0.422	0.707
			0.5	0.352	0.929	0.981	0.983
			0.7	0.922	0.983	0.983	0.982
			0.9	0.982	0.983	0.983	0.983

Table 11: The power for comparisons with a control.

N_1	N_2	N_3	θ	$r = 0.25$	$r = 0.50$	$r = 0.75$	$r = 1.00$
20	20	20	0.1	0.005	0.007	0.012	0.017
			0.3	0.084	0.387	0.740	0.927
			0.5	0.674	0.995	1.000	1.000
			0.7	0.994	1.000	1.000	1.000
			0.9	1.000	1.000	1.000	1.000
30	20	10	0.1	0.002	0.004	0.007	0.010
			0.3	0.058	0.305	0.620	0.837
			0.5	0.551	0.969	0.999	1.000
			0.7	0.966	1.000	1.000	1.000
			0.9	1.000	1.000	1.000	1.000
40	10	10	0.1	0.002	0.002	0.004	0.007
			0.3	0.033	0.202	0.507	0.778
			0.5	0.434	0.962	0.999	1.000
			0.7	0.955	1.000	1.000	1.000
			0.9	1.000	1.000	1.000	1.000

4. Concluding remarks and recommendations

The paper provides an extension of the results by Hyodo *et al.* (2013) and investigation of the robustness of the extended multiple comparison procedures under non-normality. Like those of Hyodo *et al.* (2013) our procedures are built upon the Dempster trace criterion and focus on high-dimensional cases that allow $N_i > p$. Unlike Hyodo *et al.* (2013), we examine the unbalanced case and show that consistency and asymptotic normality hold in the adjusted high-dimensional asymptotic framework. We suspect the asymptotic distribution of terms like D_{ij} in (1) and D_{1j} in (2) to be difficult to derive under the non-normality assumption and leave this question for future work. However, our simulations indicate that the extended procedures in (1) and (2) appear to perform well for a number of non-normal distributions, and hence are robust under non-normality.

We can thereby recommend the use of $D_{\max}$ -type statistics for both pairwise comparisons and comparisons with a control for the unbalanced case with very small sample sizes and very high-dimensionality. The best performance is achieved when the distribution underlying the data is not a heavy tailed one.

Acknowledgments

The authors would like to express their sincere gratitude to the referees for their helpful comments and suggestions. The research of the second author was partly supported by the Japan Society for the Promotion of Science (JSPS). The research of the third author was partly supported by a grant from the Royal Swedish Academy of Sciences, G.S. Magnuson's fund. The research of the fourth author was partly supported by a grant from the Swedish Research Council (421-2008-1966).

REFERENCES

- Dempster, A. P. (1958). A high dimensional two samples significance test. *Annals of Mathematical Statistics* **29**, 995–1010.
- Dempster, A. P. (1960). A significance test for separation of two highly multivariate small samples. *Biometrics* **16**, 41–50.
- Fujikoshi, Y., Himeno, T. and Wakaki, H. (2004). Asymptotic results of a high dimensional MANOVA test and power comparison when the dimension is large compared to the sample size. *Journal of the Japan Statistical Society* **34**, 19–26.
- Himeno, T. (2007). Asymptotic expansions of the null distributions for the Dempster trace criterion. *Hiroshima Mathematical Journal* **37**, 431–454.
- Hyodo, M., Takahashi, S. and Nishiyama, T. (2013). Multiple comparisons among mean vectors when the dimension is larger than the total sample size. *to appear in Communications in Statistics–Simulation and Computation*.
- Nishiyama, T., Hyodo, M., Seo, T. and Pavlenko, T. (2013). Testing linear hypotheses of mean vectors for high-dimension data with unequal covariance matrices. *Journal of Statistical Planning and Inference* **143**, 1898–1911.
- Seo, T. (1995). Simultaneous confidence procedures for multiple comparisons of mean vectors in multivariate normal populations. *Hiroshima Mathematical Journal* **25**, 387–422.
- Seo, T. and Siotani, M. (1992). The multivariate studentized range and its upper percentiles. *Journal of the Japan Statistical Society* **22**, 123–137.

(Received: November 26, 2012, Accepted: September 26, 2013)