AStA Advances in Statistical Analysis (2019) 103:593–618
https://doi.org/10.1007/s10182-018-00343-z

ORIGINAL PAPER

A unified approach to testing mean vectors with large dimensions

M. Rauf Ahmad[1]

Received: 19 May 2017 / Accepted: 29 November 2018 / Published online: 10 December 2018

Abstract
A unified testing framework is presented for large-dimensional mean vectors of one or several populations which may be non-normal with unequal covariance matrices. Beginning with one-sample case, the construction of tests, underlying assumptions and asymptotic theory, is systematically extended to multi-sample case. Tests are defined in terms of U-statistics-based consistent estimators, and their limits are derived under a few mild assumptions. Accuracy of the tests is shown through simulations. Real data applications, including a five-sample unbalanced MANOVA analysis on count data, are also given.

Keyword High-dimensional inference · Behrens–Fisher problem · MANOVA · U-statistics

1 Introduction

Let $\mathbf{X}_k = (X_{k1}, \ldots, X_{kp})' \sim \mathcal{F}$, $k = 1, \ldots, n$ be iid random vectors, where $\mathcal{F}$ denotes a p-variate distribution, with $E(\mathbf{X}_k) = \boldsymbol{\mu} \in \mathbb{R}^p$ and $\text{Cov}(\mathbf{X}_k) = \boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}_{>0}$. A hypothesis of foremost interest to be tested in this setup is $H_0 : \boldsymbol{\mu} = \mathbf{0}$ against an appropriate alternative, say H_1 : Not H_0. For an extension to $g \geq 2$ samples, let $\mathbf{X}_{ik} = (X_{ik1}, \ldots, X_{ikp})' \sim \mathcal{F}_i$ be iid random vectors with $E(\mathbf{X}_{ik}) = \boldsymbol{\mu}_i \in \mathbb{R}^p$, $\text{Cov}(\mathbf{X}_{ik}) = \boldsymbol{\Sigma}_i \in \mathbb{R}^{p\times p}_{>0}$, $k = 1, \ldots, n_i$, $i = 1, \ldots, g$. The corresponding hypothesis of interest is $H_{0g}: \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_g$ vs. $H_{1g}:$ Not H_{0g}.

Our objective here is to present test statistics for the aforementioned one- and multi-sample hypotheses when $p > n_i$, $\mathcal{F}_i$'s are not necessarily normal and $\boldsymbol{\Sigma}_i$, likewise n_i, in the multi-sample case may be unequal. The proposed tests are thus valid for high-dimensional, non-normal, unbalanced data under Behrens–Fisher problem.

✉ M. Rauf Ahmad
rauf.ahmad@statistik.uu.se

1 Department of Statistics, Uppsala University, Uppsala, Sweden

In particular, for $g \geq 3$, it refers to testing high-dimensional one-way MANOVA hypothesis under non-normality and multi-sample Behrens–Fisher problem.

When $p < n_i$, tests of H_0 or H_{0g} are most often carried out by Hotelling's T^2 or Wilks' Lambda statistic which are uniformly most powerful invariant likelihood ratio tests. They, however, collapse for high-dimensional case, particularly due to singularity of the empirical covariance matrix involved (see Sects. 2, 3). A number of proposals have recently been put forth in the literature on the modification of these classical tests for high-dimensional data.

Whereas most modifications assume normality, some of them are based on a more flexible model, and still others offer completely nonparametric solution to the problem. Likewise holds for homoscedasticity assumption, $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}\ \forall\ i = 1, \ldots g \geq 2$. For details, see e.g., Dempster (1958), Bai and Saranadasa (1996), Läuter et al. (1998), Läuter (2004), Fujikoshi (2004), Schott (2007), Chen and Qin (2010), Aoshima and Yata (2011, 2015), Katayama and Kano (2014), Wang et al. (2015), Feng et al. (2016) and Hu et al. (2017). For a review, see Hu and Bai (2015) and Fujikoshi et al. (2010).

We present a coherent testing theory encompassing one- and multi-sample cases. The construction of the tests, the assumptions, and the strategy of obtaining limiting distribution of the test statistics is succinctly threaded together via a common approach, initiating with the one-sample case and extending systematically to the multi-sample cases. The main distinguishing feature of the proposed tests is that we simultaneously relax commonly adopted linear model assumptions such as normality and homoscedasticity, for all cases up to one-way MANOVA. Further, all tests are defined in terms of U-statistics with simple, bivariate, product kernels composed of bilinear forms of independent vectors. This helps us determine the limits of the test statistics for a general multivariate model. These limits are derived under (n_i, p)- or high-dimensional asymptotics, i.e., $n_i, p \to \infty$, using only a few mild assumptions.

The basic idea is introduced in detail for one-sample case in the next section, with an extension to two-sample case. Multi-sample extension follows in Sect. 3. Sections 4 and 5 deal with simulations and applications. Proofs and technical details are deferred to "Appendix".

2 The one- and two-sample tests

2.1 The one-sample case

For the one-sample data setup in Sect. 1, let the unbiased estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ be defined as $\overline{\mathbf{X}} = \sum_{k=1}^{n} \mathbf{X}_k/n$ and $\widehat{\boldsymbol{\Sigma}} = \sum_{k=1}^{n}(\mathbf{X}_k - \overline{\mathbf{X}})(\mathbf{X}_k - \overline{\mathbf{X}})'/(n-1)$. If $n > p$ and $\mathcal{F}$ is multivariate normal, then $H_0 : \boldsymbol{\mu} = \mathbf{0}$ can be tested using Hotelling's statistic $T^2 = n\overline{\mathbf{X}}'\widehat{\boldsymbol{\Sigma}}^{-1}\overline{\mathbf{X}} = \overline{\mathbf{X}}'[\widehat{\boldsymbol{\Sigma}}/n]^{-1}\overline{\mathbf{X}}$ where $\widehat{\boldsymbol{\Sigma}}/n$ estimates $\boldsymbol{\Sigma}/n = \text{Cov}(\overline{\mathbf{X}})$. When $p > n$, $\widehat{\boldsymbol{\Sigma}}$ is singular and T^2 collapses, requiring a careful modification that can provide valid inference when $p \to \infty$, possibly along with $n \to \infty$. The two indices may sometimes be assumed to grow at the same rate so that $p/n \to c \in (0, \infty)$. Alternatively, a sequential asymptotics, first letting $p \to \infty$ followed by $n \to \infty$, may be considered under which conditions like $p/n \to c$ may be dispensed with.

Note that, $\widehat{\boldsymbol{\Sigma}}$ may be ill-conditioned for $p < n$ whence $(\cdot)^{-1}$ can be replaced with Moor–Penrose inverse (see e.g., Duchesne and Francq 2015). For $p \gg n$, this approach is unreliable and inefficient. An alternative is to remove $\widehat{\boldsymbol{\Sigma}}^{-1}$ from T^2 and consider the Euclidean distance $Q = \overline{\mathbf{X}}'\overline{\mathbf{X}} = \|\overline{\mathbf{X}}\|^2$. An interesting consequence of this can be witnessed by a simple split of Q as

$$Q = \frac{1}{n^2}\sum_{k=1}^{n}\sum_{r=1}^{n}\mathbf{X}_k'\mathbf{X}_r \;=\; \frac{1}{n^2}\sum_{k=1}^{n}\mathbf{X}_k'\mathbf{X}_k + \frac{1}{n^2}\sum_{\substack{k=1\\k\neq r}}^{n}\sum_{r=1}^{n}\mathbf{X}_k'\mathbf{X}_r = Q_1 + U_n, \tag{1}$$

where $Q_1 = (E - U_n)/n$ with $E = \sum_{k=1}^{n}\mathbf{X}_k'\mathbf{X}_k/n$ and $U_n = \sum_{k\neq r}^{n}\mathbf{X}_k'\mathbf{X}_r/n(n-1)$. Note that, E is an average of quadratic forms, and U_n is an average of bilinear forms composed of independent components. It is shown below that the limiting distribution of the statistic mainly follows from U_n where Q_1 converges in probability to a constant. With $E(\mathbf{X}_k'\mathbf{X}_k) = \text{tr}(\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\boldsymbol{\mu}$, $E(\mathbf{X}_k'\mathbf{X}_r) = \boldsymbol{\mu}'\boldsymbol{\mu}$, we get $E(Q_1) = \text{tr}(\boldsymbol{\Sigma})/n$, $E(U_n) = \boldsymbol{\mu}'\boldsymbol{\mu}$. Thus,

$$E(Q) = \text{tr}(\boldsymbol{\Sigma})/n + \|\boldsymbol{\mu}\|^2, \tag{2}$$

which is $\text{tr}(\boldsymbol{\Sigma})/n$ under H_0. We observe a few salient features of this bifurcation of Q. First, $E(Q_1) = \text{tr}(\boldsymbol{\Sigma})/n = \text{Cov}(\overline{\mathbf{X}})$ implies that the removal of the inverse of the estimator of $\text{Cov}(\overline{\mathbf{X}})$ results into a bias term composed of the trace of the same estimator, since it can be verified that $Q_1 = \text{tr}(\widehat{\boldsymbol{\Sigma}})/n$ or $\mathbf{E}_1 - \mathbf{Q}_0 = \widehat{\boldsymbol{\Sigma}}$ with $\mathbf{E}_1 = \sum_{k=1}^{n}\mathbf{X}_k\mathbf{X}_k'/n$ and $\mathbf{Q}_0 = \sum_{k\neq r}^{n}\mathbf{X}_k\mathbf{X}_r'/n(n-1)$ as matrix versions of E and U_n. Note also that Q_1 is independent of $\boldsymbol{\mu}$, and U_n is independent of $\boldsymbol{\Sigma}$, under both H_0 and H_1. Now, $E(U_n) = \|\boldsymbol{\mu}\|^2$ which is 0 under H_0. Together, the last two facts imply that U_n can be used to construct the modified test statistic for H_0, whereas Q_1 can help compensate for the removal of estimator from the original test statistic. For this, write $Q = \text{tr}(\widehat{\boldsymbol{\Sigma}})/n + U_n = Q_1 + U_n$ and by a simple scaling and re-writing, consider the statistic

$$T_1 = 1 + \frac{nQ_0}{nQ_1/p}, \tag{3}$$

where $Q_0 = U_n/p$ is U_n, but with kernel normed by p, $h(\mathbf{x}_k, \mathbf{x}_r) = \mathbf{X}_k'\mathbf{X}_r/p$. T_1 is the proposed modified statistic for $H_0 : \boldsymbol{\mu} = \mathbf{0}$ when $p \gg n$ and $\mathcal{F}$ may be non-normal.

For the limit of T_1, nQ_1/p is first shown to converge in probability to a constant as $n, p \to \infty$. Then, nQ_0 is shown to converge weakly to a normal limit. Under H_0, the kernel of U_n degenerates so that the null limit follows through a weighted sum of independent χ^2 variables. The limit of T follows then by Slutsky's lemma. As the same scheme will later be extended for $g \geq 2$, we treat the one-sample case in detail. Let λ_s, $s = 1, \ldots, p$ be the eigenvalues of $\boldsymbol{\Sigma}$ so that $\lambda_s/p = \nu_s$ corresponds to $\boldsymbol{\Sigma}/p$. We need the following assumptions.

Assumption 1 $E(X_{ks}^4) = \gamma_s \leq \gamma < \infty \forall\, s = 1, \ldots, p$, $\gamma_0 \in \mathbb{R}^+$.

Assumption 2 $\lim_{p\to\infty}\sum_{s=1}^{p}\nu_s = \nu_0 \in \mathbb{R}^+$.

Assumption 3 $\lim_{n,p\to\infty} p/n = c = O(1)$.

Assumption 4 $\lim_{p\to\infty} \boldsymbol{\mu}'\boldsymbol{\Sigma}\boldsymbol{\mu}/p = \phi = O(1)$.

Assumption 1 helps us relax normality. By Assumption 2, $\sum_{s=1}^{p} \nu_s^2 = O(1)$, as $p \to \infty$. Assumption 4 is only required under H_1. We have the following theorem, proved in "Appendix B.1."

Theorem 5 *For T_1 in Eq. (3), $(T_1 - E(T_1))/\sqrt{\mathrm{Var}(T_1)} \xrightarrow{\mathcal{D}} N(0, 1)$ under Assumptions 1–4, as $n, p \to \infty$, where $E(T_1)$ and $\mathrm{Var}(T_1)$ denote the mean and variance of T_1.*

From the proof of Theorem 5, $E(T_1)$ and $\mathrm{Var}(T_1)$ approximate 1 and $2\,\mathrm{tr}(\boldsymbol{\Xi}^2)/[\mathrm{tr}(\boldsymbol{\Xi})]^2$, respectively, $\boldsymbol{\Xi} = n\boldsymbol{\Sigma}/p$. As the limit follows from a weighted sum of χ_1^2 variables, the moments in fact approximate a scaled Chi-square variable, say χ_f^2/f with moments 1 and $2/f$, where $f = f_1/f_2$, $f_1 = [\mathrm{tr}(\boldsymbol{\Xi})]^2$, $f_2 = \mathrm{tr}(\boldsymbol{\Xi}^2)$. Thus, to estimate $\mathrm{Var}(T_1)$, we need consistent estimators of $\mathrm{tr}(\boldsymbol{\Sigma}^2)$ and $[\mathrm{tr}(\boldsymbol{\Sigma})]^2$. Define $Q = \sum_{k=1}^{n} (\widetilde{\mathbf{X}}_k'\widetilde{\mathbf{X}}_k)^2/(n-1)$, $\widetilde{\mathbf{X}} = \mathbf{X}_i - \overline{\mathbf{X}}$, $\eta = (n-1)/[n(n-2)(n-3)]$. Then, $E_2 = \eta\{(n-1)(n-2)\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}^2) + [\mathrm{tr}(\widehat{\boldsymbol{\Sigma}})]^2 - nQ\}$, $E_3 = \eta\{2\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}^2) + (n^2 - 3n + 1)[\mathrm{tr}(\widehat{\boldsymbol{\Sigma}})]^2 - nQ\}$ are unbiased and consistent estimators of $\mathrm{tr}(\boldsymbol{\Sigma}^2)$ and $[\mathrm{tr}(\boldsymbol{\Sigma})]^2$. Then $\widehat{f} = \widehat{f_1}/\widehat{f_2}$ is consistent estimator of f, hence $\widehat{\mathrm{Var}(T_1)}$ of $\mathrm{Var}(T_1)$ such that $\widehat{\mathrm{Var}(T_1)}/\mathrm{Var}(T_1) \to 1$; see Ahmad (2017b) and end of Sect. 3. We have the following corollary.

Corollary 6 *Theorem 5 remains valid when $\mathrm{Var}(T_1)$ is replaced with $\widehat{\mathrm{Var}}(T_1)$.*

Power of T_1 Let z_α be 100α%th quantile of $N(0, 1)$, $\beta(\boldsymbol{\theta})$ the power function of T_1 with $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ or $\boldsymbol{\theta} \in \boldsymbol{\Theta}_1$ where $\boldsymbol{\Theta}_0 = \{\mathbf{0}\}$, $\boldsymbol{\Theta}_1 = \boldsymbol{\Theta}\backslash\{\mathbf{0}\}$ are respective parameter spaces under H_0, H_1 with $\boldsymbol{\Theta} = \boldsymbol{\Theta} \cup \boldsymbol{\Theta}_1$, $\boldsymbol{\Theta}_0 \cap \boldsymbol{\Theta}_1 = \phi$. By Theorem 5, $\beta(\boldsymbol{\theta}) = \mathrm{P}(z_1 \geq z_\alpha)$ with $\beta(\boldsymbol{\theta}|H_0) = \alpha$, $\beta(\boldsymbol{\theta}|H_1) = 1 - \beta$, as $n, p \to \infty$, where $z_1 = (T_1 - E(T_1))/\sqrt{\mathrm{Var}(T_1)}$. Then, $1 - \beta = P(z \geq z_\alpha - n\delta)$, $\delta = \delta_1/\delta_2$, $\delta_1 = \boldsymbol{\mu}'\boldsymbol{\mu}/p$, $\delta_2^2 = \sum_{s=1}^{p} \nu_s^2$. By the convergence of nQ_1/p, and as δ_1, δ_2 are uniformly bounded under the assumptions, $1 - \beta \to 1$ as $n, p \to \infty$.

Remark 7 A remark on the structure of T_1 is in order. With $[nQ_1/p]/\mathrm{tr}(\boldsymbol{\Xi})$ converging in probability to 1, consider $T_1 = 1 + nU_n/\mathrm{tr}(\boldsymbol{\Sigma})$, also ignoring p for convenience. Then, $E(T_1) = 1 + n\|\boldsymbol{\mu}\|^2/\mathrm{tr}(\boldsymbol{\Sigma}) = 1 + E(\overline{\mathbf{X}})'E(\overline{\mathbf{X}})/\mathrm{Cov}(\overline{\mathbf{X}})$, where $E(T_1) = 1$ under H_0. In this sense, T_1 is similar to an F-statistic, where T_1 is close to 1 under H_0 and moves apart as $\boldsymbol{\mu}$ deviates from $\mathbf{0}$. Since $\mathrm{Cov}(\overline{\mathbf{X}}) + E(\overline{\mathbf{X}})'E(\overline{\mathbf{X}}) = E(\overline{\mathbf{X}}'\overline{\mathbf{X}})$, the partitioning used to define T_1 helps not only adjust for bias term but also makes the resulting statistic computationally much simpler, particularly under non-normality. A similar argument holds for multi-sample tests presented in the next sections.

2.2 The two-sample case

For the multi-sample setup in Sect. 1, let $g = 2$. We are interested to test $H_{02} : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ versus H_{12} : Not H_{02}. Let $\overline{\mathbf{X}}_i = \sum_{k=1}^{n_i} \mathbf{X}_{ik}/n_i$ and $\widehat{\boldsymbol{\Sigma}}_i = \sum_{k=1}^{n_i}(\mathbf{X}_{ik} - \overline{\mathbf{X}}_i)(\mathbf{X}_{ik} - \overline{\mathbf{X}}_i)'/(n_i - 1)$ be unbiased estimators of $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$. Denote $n = n_1 + n_2$. Assuming normality, $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} \forall\ i$ and $n - 2 > p$, H_{02} is usually tested by two-sample T^2,

$T^2 = [n_1n_2/n](\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2)'\widehat{\mathbf{\Sigma}}^{-1}(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2)$, where $\widehat{\mathbf{\Sigma}} = \sum_{i=1}^{2}(n_i - 1)\widehat{\mathbf{\Sigma}}_i / \sum_{i=1}^{2}(n_i - 1)$ is an estimator of common $\mathbf{\Sigma}$. For $p > n - 2$ or more generally for $p > n_i$, T^2 is invalid by the same token as for its one-sample counterpart. We consider a likewise partition of $Q = \|\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2\|^2 = \overline{\mathbf{X}}_1'\overline{\mathbf{X}}_1 + \overline{\mathbf{X}}_2'\overline{\mathbf{X}}_2 - 2\overline{\mathbf{X}}_1'\overline{\mathbf{X}}_2$ as

$$Q = \sum_{i=1}^{2} \frac{1}{n_i^2} \sum_{k=1}^{n_i}\sum_{r=1}^{n_i} \mathbf{X}_{ik}'\mathbf{X}_{ir} - \frac{2}{n_1n_2}\sum_{k=1}^{n_1}\sum_{l=1}^{n_2} \mathbf{X}_{1k}'\mathbf{X}_{2l} \; = \; Q_1 + U_0 \tag{4}$$

with $Q_1 = \sum_{i=1}^{2} Q_{i1} = \sum_{i=1}^{2} \text{tr}(\widehat{\mathbf{\Sigma}}_i)/n_i = \text{tr}(\widehat{\mathbf{\Sigma}}_0)$, $Q_{i1} = (E_i - U_{n_i})/n_i = \text{tr}(\widehat{\mathbf{\Sigma}}_i)/n_i$, $U_0 = \sum_{i=1}^{2} U_{n_i} - 2U_{n_1n_2}$, where $E_i = \sum_{k=1}^{n_i} \mathbf{X}_{ik}'\mathbf{X}_{ik}/n_i$ and

$$U_{n_i} = \frac{1}{n_i(n_i - 1)}\sum_{\substack{k=1 \\ k\neq r}}^{n_i}\sum_{r=1}^{n_i}\mathbf{X}_{ik}'\mathbf{X}_{ir}, \quad U_{n_1n_2} = \frac{1}{n_1n_2}\sum_{k=1}^{n_1}\sum_{l=1}^{n_2}\mathbf{X}_{1k}'\mathbf{X}_{2l}, \tag{5}$$

are one- and two-sample U-statistics, respectively, with symmetric kernels as bilinear forms of independent vectors. As in the one-sample case, $E(Q_{i1}) = \text{tr}(\mathbf{\Sigma}_i)/n_i \Rightarrow E(Q_1) = \text{tr}(\mathbf{\Sigma}_0)$, $\mathbf{\Sigma}_0 = \sum_{i=1}^{2} \mathbf{\Sigma}_i/n_i$ and $E(U_0) = \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2$ which vanishes under H_{02}. Thus,

$$E(Q) = \text{tr}(\mathbf{\Sigma}_0) + \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2 \; = \; \text{tr}(\mathbf{\Sigma}_0) \text{ under } H_{02}. \tag{6}$$

Again, $E(Q_1)$ is independent of $\boldsymbol{\mu}_i$, and $E(U_0)$ is independent of $\mathbf{\Sigma}_i$, under H_{02} and H_{12}. Further, $E(Q_1) = \text{tr}(\mathbf{\Sigma}_0)$, $\mathbf{\Sigma}_0 = \text{Cov}(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2)$. We thus extend T_1 in Eq. (3) for H_{02} as

$$T_2 = 1 + \frac{nQ_0}{[nQ_1/p]}, \tag{7}$$

where $Q_0 = U_0/p$ is U_0 with kernels of U_{n_i} and $U_{n_1n_2}$ scaled by p, i.e., $h(\mathbf{x}_k, \mathbf{x}_r) = \mathbf{X}_{ik}'\mathbf{X}_{ir}/p$ and $h(\mathbf{x}_k, \mathbf{x}_l) = \mathbf{X}_{1k}'\mathbf{X}_{2l}/p$, respectively. Following assumptions extend those of one-sample case, where $\nu_{is} = \lambda_{is}/p$ are eigenvalues of $\mathbf{\Xi}_i = \mathbf{\Sigma}_i/p$, $i = 1, 2$.

Assumption 8 $E(X_{iks}^4) = \gamma_{is} \leq \gamma < \infty \forall\, s = 1, \ldots, p, i = 1, \ldots, g, \gamma \in \mathbb{R}^+$.

Assumption 9 $\lim_{p\to\infty} \sum_{i=1}^{p} \nu_{is} = \sum_{s=1}^{\infty} \nu_{is} = \nu_{i0} \in \mathbb{R}^+, i = 1, \ldots, g$.

Assumption 10 $\lim_{n_i,p\to\infty} p/n_i = c_i = O(1), i = 1, \ldots, g$.

Assumption 11 $\lim_{n_i\to\infty} n/n_i = \rho_i = O(1), i = 1, \ldots, g$.

Assumption 12 $\lim_{p\to\infty} \boldsymbol{\mu}_i'\mathbf{\Sigma}_k\boldsymbol{\mu}_j/p = \phi_{ijk} \leq \phi = O(1), i, j, k = 1, \ldots, g$.

As the same assumptions will be used in Sect. 3, they are stated for $g \geq 2$. Assumption 11 is additional to those for one-sample case. It is needed to keep the limit non-degenerate when $n_i \to \infty$, $n = \sum_{i=1}^{g} n_i$. Assumption 12 is again needed only under H_{12}. Following theorem, proved in "Appendix B.2", extends Theorem 5 to two-sample case.

Theorem 13 *For T_2 in Eq. (7), $(T_2 - E(T_2))/\sqrt{\text{Var}(T_2)} \xrightarrow{\mathcal{D}} N(0,1)$ under Assumptions 8–12, as $n_i, p \to \infty$, where $E(T_2)$ and $\text{Var}(T_2)$ denote the mean and variance of T_2.*

It is interesting to see how the limit for degenerate case sums up. With ν_0 as the limit of nQ_1/p, it follows from (21) and (22) that (see e.g., Anderson et al. 1994)

$$nQ_0 \xrightarrow{\mathcal{D}} \sum_{s=1}^{\infty} \left(\sqrt{\rho_1\nu_{1s}}z_{1s} - \sqrt{\rho_2\nu_{2s}}z_{2s}\right)^2 - \nu_0$$

$$\Rightarrow T_2 \xrightarrow{\mathcal{D}} \sum_{s=1}^{\infty} \left(\sqrt{\rho_1\nu_{1s}}z_{1s} - \sqrt{\rho_2\nu_{2s}}z_{2s}\right)^2 / \nu_0,$$

with 1 and $2\sum_{s=1}^{\infty}(\rho_1\nu_{1s} - \rho_1\nu_{2s})^2/\nu_0^2$ as limiting mean and variance, where the variance approximates $2\,\text{tr}(\boldsymbol{\Xi}^2)/[\text{tr}(\boldsymbol{\Xi})]^2$, $\boldsymbol{\Xi} = n\boldsymbol{\Sigma}_0/p$, $\boldsymbol{\Sigma}_0 = \sum_{i=1}^{2}\boldsymbol{\Sigma}_i/n_i$. By the same argument of a scaled Chi-square approximation as for one-sample case, the moments correspond to those of χ^2_f/f, i.e., 1 and $2/f$, $f = f_1/f_2$, $f_1 = [\text{tr}(\boldsymbol{\Xi}_0)]^2$, $f_2 = \text{tr}(\boldsymbol{\Xi}_0^2)$. Let $E_{2i} = \eta_i\{(n_i-1)(n_i-2)\text{tr}(\widehat{\boldsymbol{\Sigma}}_i^2) + [\text{tr}(\widehat{\boldsymbol{\Sigma}}_i)]^2 - n_iQ_i\}$, $E_{3i} = \eta_i\{2\text{tr}(\widehat{\boldsymbol{\Sigma}}_i^2) + (n_i^2 - 3n_i + 1)[\text{tr}(\widehat{\boldsymbol{\Sigma}}_i)]^2 - n_iQ_i\}$, where $Q_i = \sum_{k=1}^{n_i}(\widetilde{\mathbf{X}}_{ik}'\widetilde{\mathbf{X}}_{ik})^2/(n_i-1)$, $\widetilde{\mathbf{X}}_i = \mathbf{X}_{ik} - \overline{\mathbf{X}}_i$, $\eta_i = (n_i-1)/[n_i(n_i-2)(n_i-3)]$. Further, by independence, $\text{tr}(\widehat{\boldsymbol{\Sigma}}_1\widehat{\boldsymbol{\Sigma}}_2)$ is an unbiased and consistent estimator of $\text{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)$. Plugging in f_1, f_2 leads to a consistent estimator of f, hence of $\text{Var}(T_2)$, i.e., $\widehat{\text{Var}}(T_2)$. We have the following corollary.

Corollary 14 *Theorem 13 remains valid when $\text{Var}(T_2)$ is replaced with $\widehat{\text{Var}}(T_2)$.*

Remark 15 Due to its special practical value, the two-sample test has been investigated the most, also for high-dimensional case. We briefly discuss three tests, most closely related to T_2. Denote $\kappa = n_1n_2/n$, $\omega_1 = (n-1)/(n-2)$, $\omega_2 = (n-2)^2/n(n-1)$, $n = n_1 + n_2$. Let $\xi = \|\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2\|^2 - \text{tr}(\widehat{\boldsymbol{\Sigma}})/\kappa$, where $\widehat{\boldsymbol{\Sigma}}$ is the pooled estimator of common $\boldsymbol{\Sigma}$ as given in the context of T^2 above.

Dempster (1958) proposed the first two-sample test for high-dimensional data under normality, motivated by a problem put forth by his colleagues (see Sect. 5). The test, in simpler form, is given as $T_D = \|\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2\|^2/\kappa\,\text{tr}(\widehat{\boldsymbol{\Sigma}})$. An alternative form of T_D follows by partitioning the norm in the numerator into several independent quadratic forms using an orthonormal transformation, so that the test follows an approximate F distribution with degrees of freedom estimated using a scaled Chi-square distribution. See also Dempster (1960, 1968) for details, where Bai and Saranadasa (1996) give a detailed evaluation of the approximation and power of Dempster's test.

Bai and Saranadasa (1996)'s test, $T_{\text{BS}} = \kappa\xi/\sqrt{2\omega_1}B$, is a standardization of ξ under homoscedasticity, where $B^2 = \omega_2\{\text{tr}(\widehat{\boldsymbol{\Sigma}})^2 - [\text{tr}(\widehat{\boldsymbol{\Sigma}})]^2/n\}$. Chen and Qin (2010)'s test, T_{CQ}, is a standardization of $U_0 = \sum_{i=1}^{2}U_{n_i} - 2U_{n_1n_2}$; see (4). T_{CQ} is based on the same model used for T_{BS} but relaxing normality and homoscedasticity. From the partition of Q in (4), it follows that, under the assumption of homoscedasticity, T_{D} divides the norm by the biased term, where T_{BS} and T_{CQ} subtract the same bias term from the norm, so that the numerator in both tests is U_0 with $E(U_0) = \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2 = 0$ under H_0, where for $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$, $i = 1, 2$, both tests coincide.

The proposed test, T_g, $g \geq 1$, differs from both in that it uses the removed bias term to rescale the test, where it neither requires normality nor homoscedasticity assumption. Note that, T_{CQ} is also defined without the two assumptions, but the bias adjustment, assumptions and computation of variance of the statistic are reasonably different for the two tests.

To get a more precise idea on the comparison of these tests, we did a simulation study to assess their test sizes and power. Two independent random samples of iid vectors of sizes (n_1, n_2), $n_1 \in \{10, 20, 50\}$, $n_2 = 2n_1$, each of dimension $p \in \{50, 100, 300, 500\}$, are generated from normal, t_7 and Unif[0, 1] distributions with covariance matrices, $\boldsymbol{\Sigma}_i, i = 1, 2$, compound symmetry, CS, and autoregressive of order 1, AR(1). The CS and AR(1) are defined, respectively, as $\kappa\mathbf{I} + \rho\mathbf{J}$ and $\text{Cov}(X_k, X_l) = \kappa\rho^{|k-l|}$, $\forall\, k, l$, with $\mathbf{I}$ as identity matrix and $\mathbf{J}$ a matrix of 1s. For size, we pair $\boldsymbol{\Sigma}_i$ for the two populations: both $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ CS with $\rho = 0.5$ and $\rho = 0.8$, respectively; $\boldsymbol{\Sigma}_1$ as CS, $\boldsymbol{\Sigma}_2$ as AR(1), both with $\rho = 0.5$. For power, we use CS with $\rho = 0.4$ and 0.8. We take $\kappa = 1$ for all cases. For brevity, power results are only reported for $p = 100$, for normal and t distributions.

Table 1 reports estimated test sizes of T_2, T_{BS} and T_{CQ} for all distributions with both pairs of $\boldsymbol{\Sigma}_i$. We observe an accurate performance of T_2 for all parameters, whereas T_{BS} and T_{CQ} prove, respectively, to be very liberal and very conservative, with their performance at least not improving with increasing p or (particularly) increasing n. Note that, the inaccuracy of T_{BS} can be justified as it may pertain to the homoscedasticity assumption the test is based on and which is violated in the simulations. The performance of T_{CQ}, on the other hand, can be ascribed to its assumptions, particularly on the vanishing of trace ratios such as $\text{tr}(\boldsymbol{\Sigma}^4)/[\text{tr}(\boldsymbol{\Sigma}^2)]^2$, $\text{tr}(\boldsymbol{\Sigma}^2)/[\text{tr}(\boldsymbol{\Sigma})]^2$ and $\text{tr}(\boldsymbol{\Sigma}^3)/\,\text{tr}(\boldsymbol{\Sigma})\,\text{tr}(\boldsymbol{\Sigma}^2)$, which are not satisfied for certain covariance structures, e.g., compound symmetric. A discussion on T_g is adjourned for Sect. 4, where it is evaluated in more detail.

From Fig. 1, we also observe power of T_2 higher than its competitors where the curves come closer with increasing non-centrality parameter as well as with increasing sample sizes, and this phenomenon is very similar for both distributions. Generally, a similar comparative performance and effect of sample sizes are observed for different p values; hence, not all are reported here.

3 Multi-sample test: one-way MANOVA

Here, we extend T_2 to the general case, $g \geq 2$. As usual, $\overline{\mathbf{X}}_i$ and $\widehat{\boldsymbol{\Sigma}}_i$ are unbiased estimators of $\boldsymbol{\mu}_i$, $\boldsymbol{\Sigma}_i$ $i = 1, \ldots, g$. Recall T_2 in (7) as a modification of T^2 using the Euclidean distance $\|\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2\|^2$. For H_{0g}, we sum over all pairwise norms, $\sum_{i<j} \|\overline{\mathbf{X}}_i - \overline{\mathbf{X}}_j\|^2 = \sum_{i<j}(E_i - U_{n_i})/n_i + \sum_{i<j}(U_{n_i} + U_{n_j} - 2U_{n_in_j}) = (g-1)\sum_{i=1}^g \text{tr}(\widehat{\boldsymbol{\Sigma}}_i)/n_i + (g-1)\sum_{i=1}^g U_{n_i} - 2\sum_{i<j}^g U_{n_in_j}$, and define the MANOVA statistic as

$$T_g = (g-1) + \frac{nQ_0}{[nQ_1/p]}, \tag{8}$$

$Q_1 = \sum_{i=1}^g Q_{i1}$, $Q_{i1} = (E_i - U_{n_i})/n_i = \text{tr}(\widehat{\boldsymbol{\Sigma}}_i)/n_i$, $E_i = \sum_{k=1}^{n_i} \mathbf{X}_{ik}'\mathbf{X}_{ik}/n_i$, $Q_0 = \sum_{i<j} Q_{0ij}$, $Q_{0ij} = U_{n_i} + U_{n_j} - 2U_{n_in_j}$, where U_{n_i}, $U_{n_in_j}$ are as defined in (5) with

Table 1 Estimated test size for T_2, T_{BS} and T_{CQ} for three distributions with unequal covariance matrices

$\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$	n_1, n_2	T	ND p				TD p				UD p			
			50	100	300	500	50	100	300	500	50	100	300	500
CS, CS	10, 20	T_2	0.063	0.054	0.057	0.047	0.056	0.046	0.042	0.053	0.062	0.055	0.057	0.058
		T_{BS}	0.104	0.113	0.112	0.105	0.094	0.087	0.095	0.099	0.115	0.107	0.114	0.106
		T_{CQ}	0.023	0.024	0.027	0.012	0.025	0.023	0.019	0.014	0.025	0.022	0.019	0.024
	20, 40	T_2	0.061	0.051	0.059	0.056	0.047	0.044	0.052	0.053	0.069	0.058	0.053	0.051
		T_{BS}	0.110	0.106	0.115	0.087	0.086	0.086	0.108	0.096	0.105	0.103	0.097	0.118
		T_{CQ}	0.030	0.016	0.025	0.022	0.020	0.015	0.030	0.019	0.020	0.026	0.017	0.027
	50, 100	T_2	0.064	0.052	0.066	0.053	0.049	0.054	0.057	0.055	0.054	0.055	0.058	0.052
		T_{BS}	0.104	0.097	0.107	0.099	0.091	0.096	0.089	0.104	0.083	0.107	0.100	0.118
		T_{CQ}	0.026	0.025	0.022	0.021	0.022	0.020	0.025	0.021	0.018	0.024	0.020	0.023
CS, AR	10, 20	T_2	0.043	0.048	0.056	0.044	0.036	0.052	0.036	0.056	0.046	0.054	0.054	0.059
		T_{BS}	0.118	0.181	0.160	0.135	0.103	0.140	0.119	0.144	0.104	0.135	0.150	0.163
		T_{CQ}	0.010	0.024	0.015	0.013	0.013	0.016	0.014	0.015	0.016	0.014	0.015	0.013
	20, 40	T_2	0.043	0.050	0.059	0.055	0.057	0.053	0.041	0.054	0.060	0.044	0.051	0.056
		T_{BS}	0.125	0.143	0.148	0.163	0.110	0.122	0.119	0.133	0.119	0.122	0.144	0.147
		T_{CQ}	0.012	0.012	0.014	0.021	0.016	0.011	0.013	0.012	0.015	0.013	0.013	0.011
	50, 100	T_2	0.055	0.048	0.056	0.056	0.063	0.058	0.050	0.057	0.046	0.051	0.049	0.051
		T_{BS}	0.120	0.127	0.150	0.164	0.124	0.122	0.122	0.141	0.111	0.106	0.148	0.133
		T_{CQ}	0.011	0.015	0.009	0.016	0.020	0.010	0.012	0.010	0.013	0.020	0.016	0.017

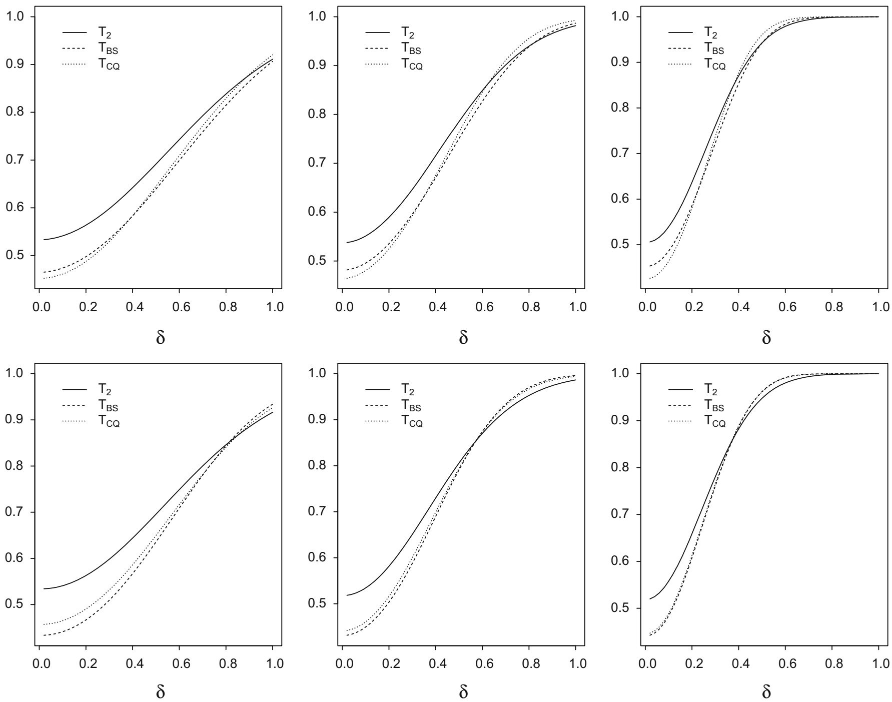

Fig. 1 Power curves of T_2, T_{BS} and T_{CQ} for normal (upper) and t (lower) distributions with (L to R) $(n_1, n_2) = (10, 20), (20, 40), (50, 100)$, $p = 100$ and CS structures with $\rho = 0.4$ and 0.8

kernels $h(\mathbf{x}_{ik}, \mathbf{x}_{ir}) = \mathbf{X}_{ik}'\mathbf{X}_{ir}, k \neq r, h(\mathbf{x}_{ik}, \mathbf{x}_{jl}) = \mathbf{X}_{ik}'\mathbf{X}_{jl}, i \neq j, k, r, l = 1, \ldots, n_i,$ $i, j = 1, \ldots, g$, $n = \sum_{i=1}^{g} n_i$. Further, $Q_1 = \operatorname{tr}(\widehat{\boldsymbol{\Sigma}}_0)$, $\widehat{\boldsymbol{\Sigma}}_0 = \sum_{i=1}^{g} \widehat{\boldsymbol{\Sigma}}_i/n_i$, is an unbiased estimator of $\operatorname{tr}(\boldsymbol{\Sigma}_0)$, $\boldsymbol{\Sigma}_0 = \sum_{i=1}^{g} \boldsymbol{\Sigma}_i/n_i$. We begin with the moments of Q_0. In particular,

$$\begin{aligned}
\operatorname{Var}(Q_0) &= \sum_{i=1}^{g} \sum_{\substack{j=1\\ i\neq j}}^{g} \operatorname{Var}(Q_{0ij}) + \sum_{i=1}^{g} \sum_{i'=1}^{g} \sum_{j=1}^{g} \sum_{\substack{j'=1\\ (i,j)\neq(i',j')}}^{g} \operatorname{Cov}(Q_{0ij}, Q_{0i'j'}) \\
&= (g-1)^2 \sum_{i=1}^{g} \operatorname{Var}(U_{n_i}) + 4 \sum_{i=1}^{g} \sum_{\substack{j=1\\ i<j}}^{g} \operatorname{Var}(U_{n_i n_j}) \\
&\quad + 8 \sum_{i=1}^{g} \sum_{\substack{j=1\\ i<j<j'}}^{g} \sum_{j'=1}^{g} \operatorname{Cov}(U_{n_i n_j}, U_{n_i n_{j'}}) \\
&\quad + 8 \sum_{i=1}^{g} \sum_{\substack{j=1\\ i<i'<j'}}^{g} \sum_{j'=1}^{g} \operatorname{Cov}(U_{n_i n_j}, U_{n_{i'} n_j})
\end{aligned}$$

$$+ 8\sum_{\substack{i=1 \\ i<i'<j'}}^{g}\sum_{j=1}^{g}\sum_{j'=1}^{g} \text{Cov}(U_{n_i n_j}, U_{n_{i'} n_{j'}})$$

$$-4(g-1)\sum_{\substack{i=1 \\ i<j<j'}}^{g}\sum_{j=1}^{g}\sum_{j'=1}^{g} \text{Cov}(U_{n_i}, U_{n_i n_j})$$

$$-4(g-1)\sum_{\substack{i=1 \\ i<j<j'}}^{g}\sum_{j=1}^{g}\sum_{j'=1}^{g} \text{Cov}(U_{n_j}, U_{n_i n_j})$$

$$-4(g-1)\sum_{\substack{i=1 \\ i<j<j'}}^{g}\sum_{j=1}^{g}\sum_{j'=1}^{g} \text{Cov}(U_{n_i}, U_{n_j n_{j'}})$$

where the covariances vanish when $i \neq i'$, $j \neq j'$. Denoting $\mathbf{\Sigma}_{0ij} = \mathbf{\Sigma}_i/n_i + \mathbf{\Sigma}_j/n_j$, $i < j$, and using the moments of U-statistics from Sect. A.2, we obtain

$$\text{Var}(Q_{0ij}) = \frac{2}{p^2}\,\text{tr}(\mathbf{\Sigma}_{0ij}^2) + \frac{4}{p^2}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)'\mathbf{\Sigma}_{0ij}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) \tag{9}$$

$$\text{Cov}(Q_{0ij}, Q_{0ij'}) = \frac{2}{n_i^2 p^2}\,\text{tr}(\mathbf{\Sigma}_i^2) + \frac{4}{n_i p^2}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)'\mathbf{\Sigma}_i(\boldsymbol{\mu}_i - \boldsymbol{\mu}_{j'}) \tag{10}$$

$$\text{Cov}(Q_{0ij}, Q_{0i'j}) = \frac{2}{n_j^2 p^2}\,\text{tr}(\mathbf{\Sigma}_j^2) + \frac{4}{n_j p^2}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)'\mathbf{\Sigma}_j(\boldsymbol{\mu}_{i'} - \boldsymbol{\mu}_j). \tag{11}$$

Theorem 16 summarizes the moments which reduce to those of two-sample case for $g = 2$.

Theorem 16 *For Q_0 defined above, we have*

$$E(Q_0) = \frac{1}{p}\sum_{\substack{i=1 \\ i<j}}^{g}\sum_{j=1}^{g}\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2$$

$$\text{Var}(Q_0) = \frac{1}{p^2}\left[2(g-1)^2\sum_{i=1}^{g}\frac{\text{tr}(\mathbf{\Sigma}_i^2)}{n_i^2} + 4\sum_{\substack{i=1 \\ i<j}}^{g}\sum_{j=1}^{g}\frac{\text{tr}(\mathbf{\Sigma}_i\mathbf{\Sigma}_j)}{n_i n_j}\right.$$

$$\left. + 4\sum_{\substack{i=1 \\ i<j}}^{g}\sum_{j=1}^{g}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)'\mathbf{\Sigma}_{0ij}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)\right.$$

$$+4\sum_{\substack{i=1\\i<j<j'}}^{g}\sum_{j=1}^{g}\sum_{j'=1}^{g}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)'\boldsymbol{\Sigma}_{0ij}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_{j'})\Bigg]$$

Now, consider the limit of T_g under Assumptions 8–12. By the independence of g samples, the convergence of Q_1 follows exactly as for $g = 2$ so that, as $n_i, p \to \infty$,

$$nQ_1/p \xrightarrow{\mathcal{P}} \nu_0,$$

where $\nu_0 = \sum_{i=1}^{g} \rho_i \nu_{i0} = \sum_{i=1}^{g}\sum_{s=1}^{\infty} \rho_i \nu_{is}$. For the limit of Q_0, we note, from the formulation $(g-1)\sum_{i=1}^{g} U_{n_i} - 2\sum_{i<j}^{g} U_{n_i n_j}$ and by the independence of U_{n_i}, U_{n_j}, $i \neq j$, which we need the distribution of $\mathbf{U}_N = (U_{n_i}, U_{n_i n_{i+1}}, \ldots, U_{n_i n_g})'$, $i = 1, \ldots, g-1$. Alternatively, we can consider $\mathbf{Q}_0 = (Q_{012}, \ldots, Q_{0g-1,g})'$. For $g = 2$, $\mathbf{U}_N = (U_{n_1},\ U_{n_1 n_2})'$, $\mathbf{Q}_0 = Q_{012}$. We can use either of the two options and proceed as for $g = 2$. $\mathbf{Q}_0$ is a $G \times 1$ vector, $G = g(g-1)/2$, with $\text{Cov}(\mathbf{Q}_0)$ a $G \times G$ partitioned matrix $\boldsymbol{\Lambda} = (\boldsymbol{\Lambda}_{ij}/p^2)_{i,j=1}^{G}$ where

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_{11} & \boldsymbol{\Lambda}_{12} & \ldots & \boldsymbol{\Lambda}_{1g} \\ \boldsymbol{\Lambda}_{21} & \boldsymbol{\Lambda}_{22} & \ldots & \boldsymbol{\Lambda}_{2g} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Lambda}_{g1} & \boldsymbol{\Lambda}_{12} & \ldots & \boldsymbol{\Lambda}_{gg} \end{pmatrix}. \tag{12}$$

Thus, $\boldsymbol{\Lambda}_{ii}/p^2 = \text{Cov}(\mathbf{Q}_{0i})$: $(g-i)\times(g-i)$, and $\boldsymbol{\Lambda}_{ij}/p^2 = \text{Cov}(\mathbf{Q}_{0i}, \mathbf{Q}_{0j})$: $(g-i)\times(g-j)$, $\boldsymbol{\Lambda}_{ji} = \boldsymbol{\Lambda}'_{ij}$, $i = 1, \ldots, g-1$, $j = i+1, \ldots, g$. Denote $a_i = \text{tr}(\boldsymbol{\Sigma}_i^2/n_i^2)$, $a_{0ij} = \text{tr}(\boldsymbol{\Sigma}_{0ij}^2)$. Then $\boldsymbol{\Lambda}_{ii} = 2(\oplus_{j=i+1}^{g} a_{0ij} + (\mathbf{J}-\mathbf{I})_{g-i} a_i)/p^2$, $\boldsymbol{\Lambda}_{ij} = 2(\mathbf{0}'\ \mathbf{1}'_{g-i} a_i\ \ \oplus_{j=i+2}^{g} a_j)'/p^2$, where $\mathbf{1}$ is vector of 1s, $\mathbf{J} = \mathbf{1}\mathbf{1}'$, $\mathbf{I}$ is identity matrix, $\oplus$ is Kronecker sum and $\mathbf{0}$ in $\boldsymbol{\Lambda}_{ij}$ is $(j-i-1)\times(g-j)$ with no $\mathbf{0}$ if $j-i-1 = 0$. For any i, $\boldsymbol{\Lambda}_{ii}$ has same off-diagonal element a_i with diagonal elements $a_{0ij} = \text{tr}(\boldsymbol{\Sigma}_{0ij}^2)$, $\boldsymbol{\Sigma}_{0ij} = \boldsymbol{\Sigma}_i/n_i + \boldsymbol{\Sigma}_j/n_j = \text{Cov}(\overline{\mathbf{X}}_i - \overline{\mathbf{X}}_j)$, $j = i+1$. Further, most off-diagonals in $\boldsymbol{\Lambda}_{ij}$ are 0, and the number of (rows with) zeros increases with j for every i, making $\boldsymbol{\Lambda}$ an increasingly sparse matrix.

The weak convergence holds for Q_{0ij} for any (i, j) in $\mathbf{Q}_0$, and we only need to take care of the nonzero off-diagonal elements in $\boldsymbol{\Lambda}$, i.e., a_i/p^2, which are uniformly bounded under the assumptions and same holds for Eqs. (9)–(11). The limit of $n\mathbf{Q}_0$, hence of nQ_0, follows then as of $\mathbf{U}_N$ for $g = 2$. Finally, Slutsky's lemma gives the limit of T_g. For the limit under H_{0g}, $E(\mathbf{Q}_0) = \mathbf{0}$, all covariances of U-statistics vanish (Sect. A.2) and Eqs. (9)–(11) reduce to $\text{Var}(Q_{0ij}) = 2\,\text{tr}(\boldsymbol{\Sigma}_{0ij}^2)$, $\text{Cov}(Q_{0ij}, Q_{0ij'}) = 2\,\text{tr}(\boldsymbol{\Sigma}_i^2)/n_i^2$, $\text{Cov}(Q_{0ij}, Q_{0i'j}) = 2\,\text{tr}(\boldsymbol{\Sigma}_j^2)/n_j^2$, which are independent of $\boldsymbol{\mu}_i$, so that we continue to assume $\boldsymbol{\mu}_i = \mathbf{0}\,\forall\, i$. In particular, from Theorem 16, $E(Q_0) = 0$ under

H_{0g} and

$$\mathrm{Var}(Q_0) = \frac{1}{p^2}\left[2(g-1)^2\sum_{i=1}^{g}\frac{\mathrm{tr}(\boldsymbol{\Sigma}_i^2)}{n_i^2} + 4\sum_{\substack{i=1\\i<j}}^{g}\sum_{j=1}^{g}\frac{\mathrm{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j)}{n_i n_j}\right],$$

which is $2\,\mathrm{tr}(\boldsymbol{\Sigma}_{012}^2)$ for $g = 2$; see Eq. (19). The null limit then also follows on the same lines as for $g = 2$. The following theorem generalizes Theorem 13 for $g \geq 2$ samples.

Theorem 17 *For T_g in Eq. (8), $(T_g - E(T_g))/\sqrt{\mathrm{Var}(T_g)} \xrightarrow{\mathcal{D}} N(0, 1)$ under Assumptions 8–12, as $n_i, p \to \infty$, where $E(T_g)$ and $\mathrm{Var}(T_g)$ denote the mean and variance of T_g.*

For the moments of T_g, note that the general distribution follows from the projection $\widehat{Q}_0 = \sum_{i<j}^{g}\widehat{Q}_{0ij} = \sum_{i<1}^{g}(\widehat{U}_{n_i} - 2\widehat{U}_{n_i n_j})$ of $\widetilde{Q} = \sum_{i<1}^{g}\widetilde{Q}_{0ij}\widetilde{Q}_{0ij} = Q_0 - E(Q_0)$, so that

$$E(\widehat{Q}_0) = \frac{1}{p}\sum_{\substack{i=1\\i<j}}^{g}\sum_{j=1}^{g}\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2$$

$$\mathrm{Var}(\widehat{Q}_0) = \frac{4}{p^2}\left[\sum_{\substack{i=1\\i<j}}^{g}\sum_{j=1}^{g}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)'\boldsymbol{\Sigma}_{0ij}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)\right.$$
$$\left.+4\sum_{\substack{i=1\\i<j<j'}}^{g}\sum_{j=1}^{g}\sum_{j'=1}^{g}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)'\boldsymbol{\Sigma}_{0ij}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_{j'})\right].$$

Likewise, under H_{0g}, the convergence of degenerate U_{n_i} and $U_{n_i n_j}$ gives

$$nQ_0 \xrightarrow{\mathcal{D}} \sum_{\substack{i=1\\i<j}}^{g}\sum_{j=1}^{g}\sum_{s=1}^{\infty}\left(\sqrt{\rho_i\nu_{is}}z_{is} - \sqrt{\rho_j\nu_{js}}z_{js}\right)^2$$

such that the limiting moments $E(nQ_0) = 0$ and $\mathrm{Var}(nQ_0) = 2\sum_{i<1}^{g}\sum_{s=1}^{\infty}(\rho_i\nu_{is} + \rho_j\nu_{js})^2$ approximate exact moments of Q_0 under H_{0g}. Combined with the limit of nQ_1/p, it gives

$$T_g \xrightarrow{\mathcal{D}} \frac{1}{\nu_0}\sum_{\substack{i=1\\ i<j}}^{g}\sum_{j=1}^{g}\sum_{s=1}^{\infty}\left(\sqrt{\rho_i\nu_{is}}z_{is} - \sqrt{\rho_j\nu_{js}}z_{js}\right)^2, \tag{13}$$

with $E(T_g) = g - 1$ and variance $\text{Var}(T_g) = 2\sum_{i<j}^{g}\sum_{s=1}^{\infty}(\rho_i\nu_{is} + \rho_j\nu_{js})^2/\nu_0^2$ which approximates $2\,\text{tr}(\boldsymbol{\Xi}^2)/[\text{tr}(\boldsymbol{\Xi})]^2$, $\boldsymbol{\Xi} = n\boldsymbol{\Sigma}_0/p$, $\boldsymbol{\Sigma}_0 = \sum_{i=1}^{g}\boldsymbol{\Sigma}_i/n_i$. Further $z_{ij} = \sqrt{\rho_i\nu_{is}}z_{is} - \sqrt{\rho_j\nu_{js}}z_{js}$ is a linear combination of independent $N(0, 1)$ variables, hence itself normal with mean 0, variance $\rho_i\nu_{is} + \rho_j\nu_{js}$. To estimate $\text{Var}(T_g)$, we note that the set of distinct non-zero elements in $\boldsymbol{\Lambda}$ is

$$\mathcal{S} = \left\{a_i = \text{tr}(\boldsymbol{\Sigma}_i^2),\ a_{ij} = \text{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j),\ i, j = 1, \ldots, g,\ i < j\right\}, \tag{14}$$

with cardinality $s_0 = \#\{\mathcal{S}\} = g(g + 1)/2$, i.e., for any g, we only need to estimate s_0 elements out of $G(G + 1)/2$ in order to estimate $\boldsymbol{\Lambda}$. With the estimators of $\text{tr}(\boldsymbol{\Sigma}_i^2)$, $[\text{tr}(\boldsymbol{\Sigma}_i)]^2$ and $\text{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j)$ same as given in the two-sample case, a consistent plug-in estimator of $\text{Var}(T_g)$ follows, leading to the following generalization of Corollary 14.

Corollary 18 *Theorem 17 remains valid when* $\text{Var}(T_g)$ *is replaced with* $\widehat{\text{Var}}(T_g)$.

Power of T_g For z_α as before, $\text{P}(T_g \geq z_\alpha\sqrt{\text{Var}(T_g)} + (g - 1)) = \alpha$, so that $1 - \beta = \text{P}(z_g \geq z_\alpha - n\delta)$ where, with $z_g = (T_g - E(T_g))/\sqrt{\text{Var}(T_g)}$, $\delta = \delta_1/\delta_2$, $\delta_1 = \sum_{i<j}^{g}\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2/p$, $\delta_2^2 = \text{tr}(\boldsymbol{\Xi}_0)$, $\boldsymbol{\Xi}_0 = n\boldsymbol{\Sigma}_0/p$, $\boldsymbol{\Sigma}_0 = \sum_{i=1}^{g}\boldsymbol{\Sigma}_i/n_i$. For $g = 2$, $\delta_1 = \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2/p$, $\boldsymbol{\Sigma}_0 = \sum_{i=1}^{2}\boldsymbol{\Sigma}_i/n_i$. A case of particular interest is when $\boldsymbol{\mu}_i$ are mutually orthogonal, $\boldsymbol{\mu}_i'\boldsymbol{\mu}_j = 0$, $\forall\ i < j$. The power function remains the same, now with $\delta_1 = (g - 1)\sum_{i=1}^{g}\|\boldsymbol{\mu}_i\|^2/p$ or, for $g = 2$, $\delta_1 = \|\boldsymbol{\mu}_1\|^2 + \|\boldsymbol{\mu}_2\|^2$.

Remark 19 This remark pertains to the trace estimators used to define consistent estimators of $\text{Var}(T_g)$. Consider one-sample case where E_2, E_3 as estimators of $\text{tr}(\boldsymbol{\Sigma}^2)$, $[\text{tr}(\boldsymbol{\Sigma})]^2$, given after Theorem 5, are defined as functions of $\widehat{\boldsymbol{\Sigma}}$ to keep them simple in formulation and efficient in computation. Alternatively, however, the same estimators can be defined as U-statistics which helps study their properties, particularly consistency, more conveniently. Let $\mathbf{D}_{kr} = \mathbf{X}_k - \mathbf{X}_r$, $k \neq r$ and define $A_{kr} = \mathbf{D}_{kr}'\mathbf{D}_{kr}$, $A_{krls}^2 = (\mathbf{D}_{kr}'\mathbf{D}_{ls})^2$. Then, we can equivalently write

$$E_2 = \frac{1}{P(n)}\sum_{\substack{k=1\\ \pi(k,\, r,\, l,\, s)}}^{n}\sum_{r=1}^{n}\sum_{l=1}^{n}\sum_{s=1}^{n}\frac{1}{12}B_{krls}, \quad E_3 = \frac{1}{P(n)}\sum_{\substack{k=1\\ \pi(k,\, r,\, l,\, s)}}^{n}\sum_{r=1}^{n}\sum_{l=1}^{n}\sum_{s=1}^{n}\frac{1}{12}C_{krls}.$$

where $B_{krls} = A_{krls}^2 + A_{klrs}^2 + A_{ksrl}^2$, $C_{krls} = A_{kr}A_{ls} + A_{kl}A_{rs} + A_{ks}A_{lr}$, $\pi(\cdot)$ means all indices pairwise unequal and $P(n) = n(n-1)(n-2)(n-3)$. This formulation of E_2, E_3 lends itself to be mathematically easily amenable using the theory of U-statistics. For details, see Ahmad (2016). The form extends directly to multi-sample cases by defining E_{2i}, E_{3i} for ith independent sample in the same way, with $\text{tr}(\widehat{\boldsymbol{\Sigma}}_i)\,\text{tr}(\widehat{\boldsymbol{\Sigma}}_j)$ and $\text{tr}(\widehat{\boldsymbol{\Sigma}}_i\widehat{\boldsymbol{\Sigma}}_j)$ estimating $\text{tr}(\boldsymbol{\Sigma}_i)\,\text{tr}(\boldsymbol{\Sigma}_j)$ and $\text{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j)$ as usual, where a U-statistic form of $\text{tr}(\widehat{\boldsymbol{\Sigma}})$ is $\sum_{k\neq r}^{n} A_{kr}/n(n - 1)$. For details, see Ahmad (2017a, b).

Remark 20 Note that, the Chi-square approximation in both one- and multi-sample cases follows through two-moment approximation of the limit of the test statistics with that of a scaled Chi-square variable. Box (1954a, b) used this approximation to study the violation of assumptions of homoscedastic and uncorrelated errors in ANOVA settings, later extended and modified by Geissser and Greenhouse (1958), Greenhouse and Geissser (1959) and Huynh and Feldt (1970, 1976).

4 Simulations

We evaluate the accuracy of tests for size control and power, specifically focusing on violation of normality and homoscedasticity assumptions. We take $g = 1$ and 3 and generate data from Normal, Exponential and Uniform distributions with $n = 10, 20, 50$ for T_1 and $(n_1, n_2, n_3) = (10, 15, 20), (5, 25, 50), (10, 30, 60)$, for T_3, where the last two triplets represent seriously unbalanced designs. For dimension, we take $p \in \{50, 100, 300, 500, 1000\}$. For covariances structures, we use compound symmetry (CS), autoregressive of order 1, AR(1), as defined in Sect. 2.2, and unstructured (UN), defined as $\mathbf{\Sigma} = (\sigma_{ij})_{i,j=1}^{d}$ with $\sigma_{ij} = 1(1)d \ (i = j)$, $\rho_{ij} = (i-1)/d \ (i > j)$, with $\mathbf{I}$ as identity matrix and $\mathbf{J}$ as matrix of 1s. We use $\rho = 0.5$, $\kappa = 1$.

We use $\alpha = 0.01, 0.05, 0.10$ and estimate test size by averaging $P(T \leq T_o|H_0)$ over 1000 simulations, where T denotes T_1 or T_3 and T_o is their observed value under H_0. Tables 2 and 3 report estimated size and power of T_1 for normal and exponential distributions, and Tables 4 and 5 report the same for T_3 for all distributions. For power, we fix $\alpha = 0.05$ and estimate the power by averaging $P(T \geq T_o|H_1)$ over 1000 runs, where H_1 is defined as $\boldsymbol{\mu} = \delta_r \mathbf{p}_1$, $\mathbf{p}_1 = (1/p, \ldots, p/p)$, $\delta_r = 0.2(0.2)1$. Note that, T_3 is assessed under a triplet of covariance structures (CS, AR, UN) followed by the three populations.

We observe an accurate size control for normal as well as for non-normal distributions and under all covariance structures. The stability of the size control for increasing p, for n as small as 10, is also evident. We observe a similar performance for power, with discernably better performance under AR and UN structures than under CS, for all distributions, which might be attributed to the spiky nature of CS. The power, however, also improves reasonably under CS for increasing n and p. For $g = 3$, we also observe accuracy for unbalanced design, with a drastic improvement for the last triplet of n_i. Although not reported here, similar results were observed for other ρ values in CS and AR, for other covariance structures, e.g., Toeplitz, and for other distributions, e.g., t.

We also assessed the power of proposed tests under possible sparse alternatives. For simplicity, we report results for T_1 for normal distribution with same n as used above and $p \in \{60, 100, 200\}$. We consider three levels of sparsity: small, medium and large with 25%, 50% and 75% zeros in the mean vector, respectively. Note that, 0% sparsity implies the case under H_1, where 100% sparsity implies the null case. Table 6 reports the results. Generally, the power is high under all parameter settings, indicating the validity of tests for such alternatives. Further, the power increases with increasing sample size, so that even under sparsity, the test shows a high probability to tell the null from the alternative, particularly as the sample size grows.

Table 2 Estimated size of T_1: normal and exponential distributions

n	p	CS			AR			UN		
		0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
Normal										
10	50	0.020	0.065	0.115	0.025	0.071	0.116	0.025	0.068	0.120
	100	0.019	0.069	0.127	0.024	0.080	0.136	0.020	0.069	0.113
	200	0.022	0.073	0.134	0.020	0.065	0.122	0.024	0.077	0.131
	300	0.019	0.068	0.116	0.020	0.067	0.120	0.023	0.075	0.133
	500	0.016	0.068	0.127	0.018	0.062	0.114	0.021	0.064	0.130
20	50	0.015	0.060	0.112	0.020	0.058	0.112	0.014	0.047	0.098
	100	0.011	0.053	0.098	0.014	0.054	0.109	0.015	0.055	0.109
	200	0.018	0.060	0.114	0.016	0.053	0.107	0.016	0.063	0.109
	300	0.016	0.056	0.113	0.012	0.055	0.108	0.011	0.056	0.114
	500	0.012	0.055	0.102	0.014	0.057	0.104	0.012	0.062	0.114
50	50	0.013	0.043	0.093	0.018	0.053	0.094	0.014	0.051	0.103
	100	0.015	0.048	0.102	0.011	0.044	0.089	0.014	0.052	0.107
	200	0.012	0.050	0.098	0.013	0.050	0.101	0.012	0.060	0.104
	300	0.010	0.048	0.099	0.014	0.057	0.107	0.013	0.051	0.094
	500	0.009	0.052	0.097	0.012	0.056	0.102	0.012	0.050	0.108
Exp										
10	50	0.053	0.105	0.137	0.021	0.065	0.121	0.021	0.069	0.112
	100	0.048	0.074	0.138	0.023	0.063	0.125	0.024	0.057	0.119
	300	0.033	0.068	0.125	0.018	0.065	0.113	0.016	0.062	0.108
	500	0.021	0.650	0.114	0.014	0.059	0.117	0.018	0.063	0.113
	1000	0.015	0.610	0.118	0.016	0.054	0.111	0.015	0.061	0.114
20	50	0.013	0.057	0.107	0.011	0.053	0.108	0.016	0.052	0.103
	100	0.007	0.051	0.103	0.009	0.051	0.102	0.015	0.051	0.106
	300	0.015	0.062	0.118	0.012	0.053	0.102	0.012	0.051	0.101
	500	0.011	0.049	0.102	0.013	0.054	0.110	0.010	0.056	0.110
	1000	0.012	0.058	0.110	0.010	0.052	0.095	0.012	0.056	0.113
50	50	0.013	0.057	0.102	0.011	0.052	0.110	0.009	0.049	0.101
	100	0.011	0.059	0.104	0.011	0.051	0.102	0.011	0.055	0.105
	300	0.008	0.047	0.105	0.013	0.054	0.103	0.010	0.050	0.101
	500	0.008	0.051	0.097	0.011	0.049	0.097	0.009	0.045	0.093
	1000	0.011	0.048	0.101	0.010	0.051	0.105	0.010	0.049	0.102

5 Analyses of real data sets

Figure 2 depicts average counts of macrobenthos observed along an approximately 2000 km long transact of Norwegian continental shelf. The transact under observation comprised a range of water depths and sediment properties. A total of $p = 809$ species were observed from $n = 101$ independent sites in five different regions of the transact,

Table 3 Estimated power of T_1: normal and exponential distributions

n	p	CS			AR			UN		
		0.2	0.6	1.0	0.2	0.6	1.0	0.2	0.6	1.0
Normal										
10	50	0.198	0.780	0.999	0.201	0.930	0.994	0.331	0.945	1.000
	100	0.258	0.949	1.000	0.260	0.948	1.000	0.255	0.946	1.000
	300	0.487	1.000	1.000	0.501	1.000	1.000	0.488	1.000	1.000
	500	0.650	1.000	1.000	0.666	1.000	1.000	0.643	1.000	1.000
	1000	0.839	1.000	1.000	0.805	1.000	1.000	0.858	1.000	1.000
20	50	0.397	0.998	1.000	0.384	0.998	1.000	0.393	0.995	1.000
	100	0.556	1.000	1.000	0.562	1.000	1.000	0.570	1.000	1.000
	300	0.904	1.000	1.000	0.910	1.000	1.000	0.908	1.000	1.000
	500	0.987	1.000	1.000	0.987	1.000	1.000	0.987	1.000	1.000
	1000	0.990	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
50	50	0.888	1.000	1.000	0.883	1.000	1.000	0.897	1.000	1.000
	100	0.990	1.000	1.000	0.990	1.000	1.000	0.988	1.000	1.000
	300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Exp										
10	50	0.124	0.308	0.678	0.138	0.514	0.892	0.162	0.676	0.990
	100	0.126	0.329	0.714	0.188	0.707	1.000	0.206	0.900	1.000
	300	0.201	0.413	0.778	0.350	0.907	1.000	0.462	1.000	1.000
	500	0.255	0.491	0.802	0.504	0.999	1.000	0.706	1.000	1.000
	1000	0.302	0.561	0.881	0.735	1.000	1.000	0.854	1.000	1.000
20	50	0.242	0.502	0.701	0.303	0.890	1.000	0.654	1.000	1.000
	100	0.337	0.521	0.898	0.418	0.987	1.000	0.857	1.000	1.000
	300	0.498	0.665	0.997	0.734	1.000	1.000	0.999	1.000	1.000
	500	0.605	0.717	1.000	0.871	1.000	1.000	1.000	1.000	1.000
	1000	0.723	0.815	1.000	0.912	1.000	1.000	1.000	1.000	1.000
50	50	0.458	0.748	0.998	0.682	1.000	1.000	0.898	1.000	1.000
	100	0.554	0.795	0.999	0.879	1.000	1.000	0.977	1.000	1.000
	300	0.714	0.823	1.000	0.998	1.000	1.000	1.000	1.000	1.000
	500	0.831	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1000	0.885	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

where $n_1 = 16, n_2 = 21, n_3 = 25, n_4 = 19, n_5 = 20$. Each count is a five-replicate pooled observation, and the data contain a large amount of zeros where no species could be recorded. For details, see Ellingsen and Gray (2002).

In our notation, $\mathbf{X} = (\mathbf{X}_1', \ldots, \mathbf{X}_5')' \in \mathbb{R}^{n \times p}$ represents the complete data matrix with regionwise data matrices $\mathbf{X}_i = (\mathbf{X}_{i1}', \ldots, \mathbf{X}_{in_i}')' \in \mathbb{R}^{n_i \times p}$, $\mathbf{X}_{ik} \in \mathbb{R}^p$, where n_i and p are given above. It is thus an unbalanced one-way MANOVA experiment

Table 4 Estimated size of T_3 for (CS, AR, UN) structures: all distributions

n_1, n_2, n_3	p	$N(0, 1)$			Exp(1)			Unif[0, 1]		
		0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
10, 15, 20	50	0.006	0.041	0.090	0.006	0.040	0.086	0.008	0.043	0.092
	100	0.009	0.049	0.091	0.013	0.052	0.094	0.013	0.056	0.105
	300	0.011	0.047	0.094	0.010	0.047	0.090	0.011	0.048	0.100
	500	0.010	0.047	0.093	0.009	0.045	0.089	0.012	0.052	0.098
	1000	0.011	0.048	0.096	0.010	0.046	0.098	0.011	0.052	0.097
5, 25, 50	50	0.007	0.044	0.089	0.011	0.043	0.093	0.008	0.043	0.093
	100	0.005	0.040	0.092	0.004	0.035	0.078	0.005	0.043	0.090
	300	0.004	0.034	0.084	0.004	0.034	0.084	0.004	0.038	0.084
	500	0.004	0.037	0.088	0.002	0.031	0.080	0.004	0.037	0.087
	1000	0.005	0.035	0.082	0.003	0.036	0.085	0.007	0.034	0.081
10, 30, 60	50	0.009	0.050	0.096	0.008	0.042	0.083	0.014	0.049	0.093
	100	0.011	0.051	0.096	0.009	0.044	0.087	0.014	0.055	0.101
	300	0.008	0.043	0.092	0.008	0.040	0.085	0.001	0.048	0.096
	500	0.008	0.042	0.086	0.009	0.048	0.099	0.010	0.047	0.094
	1000	0.008	0.044	0.090	0.006	0.040	0.093	0.007	0.042	0.090

Table 5 Estimated power of T_3 for (CS, AR, UN) structures: all distributions

n_1, n_2, n_3	p	$N(0, 1)$			Exp(1)			Unif[0, 1]		
		0.2	0.6	1.0	0.2	0.6	1.0	0.2	0.6	1.0
10, 15, 20	50	0.074	0.405	0.921	0.058	0.399	0.934	0.066	0.414	0.919
	100	0.071	0.598	1.000	0.064	0.603	0.995	0.078	0.600	0.993
	300	0.097	0.912	1.000	0.099	0.931	1.000	0.097	0.921	1.000
	500	0.123	0.986	1.000	0.122	0.985	1.000	0.127	0.988	1.000
	1000	0.178	0.999	1.000	0.173	0.999	1.000	0.171	1.000	1.000
5, 25, 50	50	0.109	0.887	1.000	0.117	0.899	1.000	0.101	0.871	1.000
	100	0.142	0.988	1.000	0.137	0.988	1.000	0.143	0.990	1.000
	300	0.237	1.000	1.000	0.226	1.000	1.000	0.241	1.000	1.000
	500	0.313	1.000	1.000	0.311	1.000	1.000	0.317	1.000	1.000
	1000	0.500	1.000	1.000	0.506	1.000	1.000	0.490	1.000	1.000
10, 30, 60	50	0.070	0.525	0.988	0.056	0.539	0.995	0.070	0.5418	0.988
	100	0.079	0.768	0.999	0.071	0.793	1.000	0.078	0.763	1.000
	300	0.116	0.991	1.000	0.103	0.995	1.000	0.117	0.990	1.000
	500	0.152	0.999	1.000	0.141	1.000	1.000	0.147	1.000	1.000
	1000	0.218	1.000	1.000	0.208	1.000	1.000	0.213	1.000	1.000

Table 6 Estimated power of T_1 under three sparse alternatives

n	p	25%			50%			75%		
		0.2	0.6	1.0	0.2	0.6	1.0	0.2	0.6	1.0
10	60	0.679	0.969	0.999	0.629	0.924	0.992	0.593	0.828	0.956
	100	0.680	0.973	0.999	0.643	0.938	0.995	0.600	0.839	0.963
	200	0.691	0.976	0.999	0.639	0.939	0.996	0.596	0.840	0.968
20	60	0.770	0.998	1.000	0.698	0.987	1.000	0.628	0.932	0.994
	100	0.771	0.999	1.000	0.712	0.992	1.000	0.619	0.937	0.996
	200	0.774	0.999	1.000	0.711	0.992	1.000	0.636	0.946	0.997
50	60	0.921	1.000	1.000	0.846	1.000	1.000	0.729	0.994	1.000
	100	0.933	1.000	1.000	0.866	1.000	1.000	0.737	0.997	1.000
	200	0.936	1.000	1.000	0.870	1.000	1.000	0.741	0.998	1.000

with $g = 5$ independent samples, each of n_i iid vectors of dimension 809, where $n = \sum_{i=1}^{5} n_i = 101$. The linear model can be expressed as

$$\mathbf{X}_{ik} = \boldsymbol{\mu}_i + \boldsymbol{\epsilon}_{ik}, \quad j = 1, \ldots, n_i, \; i = 1, \ldots, 5, \tag{15}$$

where the vector $\mathbf{X}_{ik}$ consists of 809 species counts measured for kth replicate (site) from ith region, $\boldsymbol{\mu}_i \in \mathbb{R}^p$ is the true average count vector of ith region, and $\boldsymbol{\epsilon}_{ik} \in \mathbb{R}^p$ are random error vectors, associated with each $\mathbf{X}_{ik}$, with $E(\boldsymbol{\epsilon}_{ik}) = \mathbf{0}$ and $\text{Cov}(\boldsymbol{\epsilon}_{ik}) = \boldsymbol{\Sigma}_i \forall\, k, i = 1, \ldots, 5$. The hypothesis of interest can be formulated as $H_{05} : \boldsymbol{\mu}_1 = \ldots = \boldsymbol{\mu}_5$ vs $H_{15} : \boldsymbol{\mu}_i \neq \boldsymbol{\mu}_j$ for at least one pair $i \neq j, i, j = 1, \ldots, 5$. We use T_g in Eq. (8) to test H_{05}.

We also apply the proposed test to two well-known data sets, referred here to as alcohol and leukemia data. The alcohol data is a two-group ($g = 2$) data that motivated Dempster to construct the first two-sample high-dimensional test (Dempster 1958); see also Dempster (1960, 1968). The data consist of $p = 59$ biochemistry measurements on $n_1 = 8$ alcoholic and $n_2 = 4$ control individuals aged 16–39 years; see also Beerstecher et al. (1950). The three-group ($g = 3$) leukemia data are often also used for classification. It consist of measurements on patients with acute lymphoblastic leukemia (ALL) carrying a chromosomal translocation involving mixed-lineage leukemia (MLL) gene. A total of $p = 11225$ gene expression profiles of leukemia cells are taken from patients in ALL group ($n_1 = 28$), B-precursor ALL carrying an MLL translocation ($n_2 = 24$) and conventional B-precursor without MLL translocation ($n_3 = 20$); see Armstrong et al. (2002) for details.

Model (15) remains the same for alcohol and leukemia data sets, with $g = 2$ and 3, respectively, and with corresponding sample sizes given above. The analyses of all three data sets are reported in Table 7. The first three columns report the data sizes and the next three the Chi-square approximation for T_g, and the penultimate two columns provide the corresponding normal approximation. Only for alcohol data, the results provide evidence in support of null hypothesis of no difference of mean vectors, whereas the hypotheses are significantly rejected for both leukemia and species data.

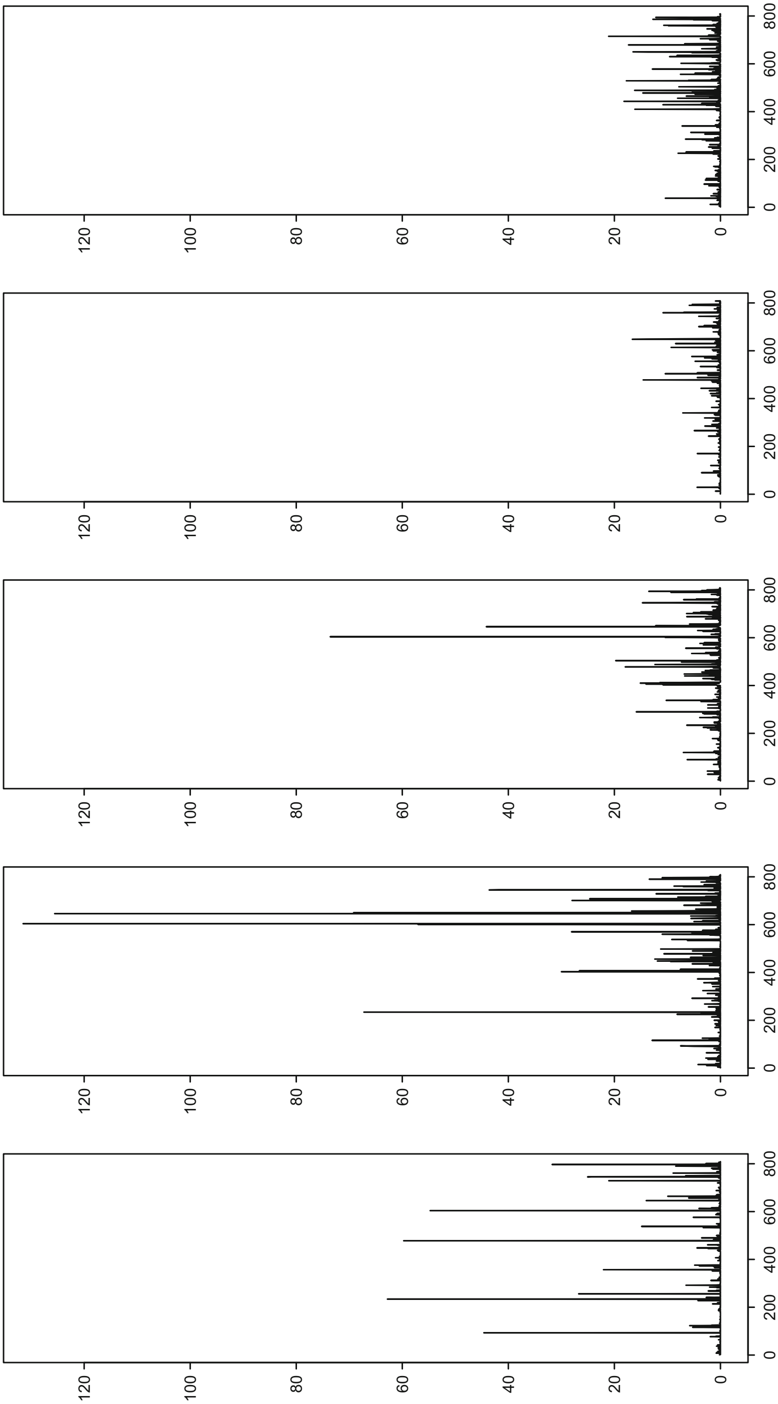

Fig. 2 Average species count of macrobenthos data for five regions

Table 7 Analysis of example data sets

Data	Data size			Chi-square test			Normal test	
	g	$(n_1, \ldots, n_g)$	p	T_g	df	p value	T_g	p value
Alcohol	2	(8, 4)	59	2.80	3.91	0.578	−0.40	0.654
Leukemia	3	(28, 24, 20)	11225	96.93	7.31	0.000	21.52	0.000
Species	5	(16, 21, 25, 19, 20)	809	180.40	7.03	0.000	40.61	0.000

The conclusions for all three data sets are consistent for both approximations. In particular, the results for species data substantiate what can be roughly witnessed in Fig. 2.

6 Discussion and remarks

Test statistics for high-dimensional mean vectors are presented. A unified strategy is proposed that systematically encompasses one- and multi-sample cases. The tests are constructed as linear combinations of U-statistics-based estimators and are valid for any distribution with finite fourth moment. The limiting distributions of the tests are derived under a few mild assumptions. Simulations are used to show the accuracy of the tests for moderate sample size and any large dimension. The tests are location invariant, so that the mean vectors need not be assumed zero. Due to singularity of empirical covariance matrix in high-dimensional case, an affine-invariant test is not possible, and location-invariance is the best that can be achieved in this case.

Acknowledgements The author is thankful to the editor, the associate editor and anonymous referee for their constructive comments which helped improve the original version of the manuscript. Kari Ellingsen's kind permission to use species data is also duly acknowledged.

A Some miscellaneous results

A.1 *U*-statistics

First, we need to set some notations. For details, see e.g., Serfling (1980), Koroljuk and Borovskich (1994), van der Vaart (1998) and Lehmann (1999). For iid X_i, let $h(X_1, \ldots, X_m) : \mathbb{R}^m \to \mathbb{R}$ denote the kernel of an mth order U-statistic, U_n, with $E(U_n) = \theta = E[h(\cdot)]$ with its projection $h_c(x_1, \ldots, x_c) = E[h(\cdot)|x_1, \ldots, x_c]$, $h_m(\cdot) = h(\cdot)$ and $\xi_c = \text{Var}[h_c(\cdot)$, $c = 1, \ldots, m$, so that $\text{Var}(U_n) = \sum_{c=1}^{m} \binom{m}{c}\binom{n-m}{m-c}\xi_c / \binom{n}{m}$. If $0 < \xi_c < \infty \forall\ c$, then $(U_n - E(U_n))/\sqrt{\text{Var}(U_n)} \xrightarrow{\mathcal{D}} N(0, 1)$. For two U-statistics, U_{ni}, of order m_i, kernels

$h_i(\cdot)$, projections $h_{ic}(\cdot)$, $i = 1, 2$, let $\xi_{cc} = \text{Cov}[h_{1c}(\cdot), h_{2c}(\cdot)]$, $c = 1, \ldots, m_1 \leq m_2$. Then, $\text{Cov}(U_{n1}, U_{n2}) = \sum_{c=1}^{m_1} \binom{m_2}{c}\binom{n-m_2}{m_1-c}\xi_{cc}/\binom{n}{m_1}$. Let $U_{n_1n_2}$ be a U-statistic of two independent samples, with kernel $h(X_{11}, \ldots, X_{1m_1}, X_{21}, \ldots, X_{2m_2})$, symmetric in each sample, projection $h_{c_1c_2} = E[h(\cdot)|X_{11}, \ldots, X_{1c_1}; X_{21}, \ldots, X_{2c_2}]$, $\xi_{c_1c_2} = \text{Cov}[h(\cdot), h_{c_1c_2}(\cdot)]$, $\xi_{00} = 0$, $c_i = 0, 1, \ldots, m_i$. Then, $\text{Var}(U_{n_1n_2}) = \sum_{c_1=0}^{m_1}\sum_{c_2=0}^{m_2}\binom{m_1}{c_1}\binom{n_1-m_1}{m_1-c_1}\binom{m_2}{c_2}\binom{n_2-m_2}{m_2-c_2}\xi_{c_1c_2}/\binom{n_1}{m_1}\binom{n_2}{m_2}$. If $0 \leq n_i/n \leq 1, n = n_1+n_2$, $0 < \xi_{c_1c_2} < \infty \; \forall \, c_i$, then $(U_{n_1n_2} - E(U_{n_1n_2}))/\sqrt{\text{Var}(U_{n_1n_2})} \xrightarrow{\mathcal{D}} N(0, 1)$.

Lemma 21 (Jiang 2010, p. 183; Hájek et al. 1999, p. 184) *Let $Y_1, Y_2, \ldots$ be iid random variables, $E(Y_i) = 0$, $\text{Var}(Y_i) = 1$. Let b_{ni} be a sequence of constants, $1 \leq i \leq n$. Then $\sum_{i=1}^{n} b_{ni}Y_i \xrightarrow{\mathcal{D}} N(0, 1)$ given $\max_i b_{ni}^2 \to 0$, as $n \to \infty$.*

A.2 Basic moments of *U*-statistics

For U_{n_i}, $h(\mathbf{X}_{ik}, \mathbf{X}_{ir}) = \mathbf{X}_{ik}'\mathbf{X}_{ir}$, $m = 2$, $h_1(\mathbf{X}_{ik}) = \boldsymbol{\mu}_i'\mathbf{X}_{ik}$, $\xi_1 = \boldsymbol{\mu}_i'\boldsymbol{\Sigma}_i\boldsymbol{\mu}_i$, $\xi_2 = \text{tr}(\boldsymbol{\Sigma}_i^2) + 2\boldsymbol{\mu}_i'\boldsymbol{\Sigma}_i\boldsymbol{\mu}_i$. For $U_{n_in_j}$, $h(\mathbf{X}_{ik}, \mathbf{X}_{jl}) = \mathbf{X}_{ik}'\mathbf{X}_{jl}$, $m_1 = 1 = m_2$, $h_{10} = \boldsymbol{\mu}_j'\mathbf{X}_{ik}$, $h_{01} = \boldsymbol{\mu}_i'\mathbf{X}_{jl}$, $h_{11}(\cdot) = h(\cdot)$, $\xi_{10} = \boldsymbol{\mu}_j'\boldsymbol{\Sigma}_i\boldsymbol{\mu}_j$, $\xi_{01} = \boldsymbol{\mu}_i'\boldsymbol{\Sigma}_j\boldsymbol{\mu}_i$, $\xi_{11} = \boldsymbol{\mu}_i'\boldsymbol{\Sigma}_j\boldsymbol{\mu}_i + \boldsymbol{\mu}_j'\boldsymbol{\Sigma}_i\boldsymbol{\mu}_j + \text{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j)$. Then, for $i \neq j$, $i \neq j'$, $i' \neq j$, $\text{Var}(U_{n_i}) = 2[2(n_i-1)\boldsymbol{\mu}_i'\boldsymbol{\Sigma}_i\boldsymbol{\mu}_i + \text{tr}(\boldsymbol{\Sigma}_i^2)]/n_i(n_i-1)$, $\text{Var}(U_{n_in_j}) = [n_i\boldsymbol{\mu}_i'\boldsymbol{\Sigma}_j\boldsymbol{\mu}_i + n_j\boldsymbol{\mu}_j'\boldsymbol{\Sigma}_i\boldsymbol{\mu}_j + \text{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j)]/n_in_j$, $\text{Cov}(U_{n_i}, U_{n_in_j}) = 2\boldsymbol{\mu}_j'\boldsymbol{\Sigma}_i\boldsymbol{\mu}_i/n_i$, $\text{Cov}(U_{n_j}, U_{n_in_j}) = 2\boldsymbol{\mu}_i'\boldsymbol{\Sigma}_j\boldsymbol{\mu}_j/n_j$, $\text{Cov}(U_{n_in_j}, U_{n_in_{j'}}) = \boldsymbol{\mu}_j'\boldsymbol{\Sigma}_i\boldsymbol{\mu}_{j'}/n_i$, $\text{Cov}(U_{n_in_j}, U_{n_{i'}n_j}) = \boldsymbol{\mu}_i'\boldsymbol{\Sigma}_j\boldsymbol{\mu}_{i'}/n_j$. See Sect. A.1 for basic notations and general moment expressions.

B Main proofs

B.1 Proof of Theorem 5

First, $E(Q_1/p) = \text{tr}(\boldsymbol{\Sigma})/p = \sum_{s=1}^{p}\nu_s$, bounded by ν_0, under Assumption 2, as $p \to \infty$. Now

$$\text{Var}(nQ_1/p) = \text{Var}(E/p) + \text{Var}(U_n/p) - 2\,\text{Cov}(E/p, U_n/p). \tag{16}$$

With $\text{Var}(\mathbf{X}_k'\mathbf{X}_k) \leq \gamma p^2$ under Assumption 1, $\text{Var}(E/p) \leq \gamma/n = O(1/n)$. From Sect. A.2, $\text{Var}(U_n/p) = \gamma/n + 2\,\text{tr}(\boldsymbol{\Sigma}^2)/n(n-1)p^2 + 4\boldsymbol{\mu}'\boldsymbol{\Sigma}\boldsymbol{\mu}/(n-1)p^2 = O(1/n)$ under the assumptions. Finally, $\text{Cov}(E/p, U_n/p) = 0$ for $\boldsymbol{\mu} = \mathbf{0}$ which can be assumed w.o.l.o.g. since $E(Q_1/p)$ does not depend on $\boldsymbol{\mu}$. Alternatively, by Cauchy–Schwarz inequality, $\text{Cov}(E/p, U_n/p) \leq [\text{Var}(E)\,\text{Var}(U_n)]^{1/2}$ which simplifies to $O(1/n)$. This proves consistency of nQ_1/p. Note that, this consistency holds both under simultaneous and sequential (n, p)-asymptotics where in the later case the last term vanishes with p, before the limit over n is carried out. Further, the limit is the same under H_0 and H_1.

Now, consider nU_n with $h(\mathbf{x}_k, \mathbf{x}_r) = \mathbf{X}_k'\mathbf{X}_r/p$ so that $E[h(\cdot)] = \|\boldsymbol{\mu}\|^2 = E(U_n)$. Define $\widetilde{U}_n = U_n - E(U_n)$ with corresponding kernel $\widetilde{h}(\cdot) = h(\cdot) - E[h(\cdot)]$. Let $\widehat{U}_n$

denotes the projection of $\widetilde{U}_n$. As U_n is a second-order U-statistic with product kernel (bilinear form of independent components), $h(\cdot)$, following the notation in Sect. A.1,

$$\widehat{U}_n = \sum_{k=1}^{n} E(\widetilde{U}_n|\mathbf{X}_k) = \frac{2}{n}\sum_{k=1}^{n} \boldsymbol{\mu}'(\mathbf{X}_k - \boldsymbol{\mu})$$

with $E(\widehat{U}_n) = 0 = E(\widetilde{U}_n)$ and $\text{Cov}(\widehat{U}_n, \widetilde{U}_n) = 4\boldsymbol{\mu}'\boldsymbol{\Sigma}\boldsymbol{\mu}/n = \text{Var}(\widehat{U}_n)$ so that, with $\text{Var}(\widetilde{U}_n)$ as given above (see Sect. A.2), it follows that $\text{Var}(n\widehat{U}_n)$ and $\text{Var}(n\widetilde{U}_n)$ are uniformly bounded under Assumptions 2 and 4, such that $\text{Var}(n\widehat{U}_n)/\text{Var}(n\widetilde{U}_n) \to 1$; see e.g., Lehmann (1999, Ch.6), Serfling (1980, Ch.5) or van der Vaart (1998, Ch.12). This, along with the convergence of nQ_1/p, gives normal limit of $nU_n/[nQ_1/p]$, hence of T, by Slutsky's theorem.

Some remarks concerning the aforementioned limit will help us extend it further under the null. To begin with, the first-order projection of $h(\cdot)$, $h_1(\mathbf{x}_k) = E[h(\cdot)|\mathbf{x}_k] = \boldsymbol{\mu}'\mathbf{X}_k/p$, along with its variance, $\xi_1 = \text{Var}[h_1(\mathbf{x}_k)] = \boldsymbol{\mu}'\boldsymbol{\Sigma}\boldsymbol{\mu}/p^2$ exactly vanishes under $H_0 : \boldsymbol{\mu} = \mathbf{0}$, making the kernel (first-order) degenerate under H_0. Note that, for the limit under H_1 above, the term involving this projection, $4\boldsymbol{\mu}'\boldsymbol{\Sigma}\boldsymbol{\mu}/np^2$ is eventually bounded under Assumption 4, for simultaneous (n, p)-asymptotics, when used for nU_n. But under sequential asymptotics, if $p \to \infty$ first, then the projection vanishes asymptotically. But the limit under H_1 still holds since the total variance $\text{Var}(U_n)$ still remains bounded under the assumptions. In fact, an additional advantage under sequential asymptotics is that now the power of T does not depend on any specific $\boldsymbol{\mu}$.

Under H_0, however, the projection and its variance ξ_1 are exactly zero and the limit need to be derived differently. Since $E[h^2(\cdot)] = \text{tr}(\boldsymbol{\Sigma}^2)/p^2 < \infty$ under the consequence of Assumption 2, the kernel is square integrable. As we shall see in the sequel, $h(\cdot)$ being a product kernel makes it further convenient to derive the limit. Without loss of generality, we can assume that the data $\mathbf{X}_k$ are generated by a separable (Hilbert) space $\mathcal{L}_2(\mathcal{X}, \mathcal{A}, \mathcal{P})$. By symmetry and square integrability of $h(\cdot)$, the map $T : \mathcal{L}_2(\mathcal{X}, \mathcal{A}, \mathcal{P}) \to \mathcal{L}_2(\mathcal{X}, \mathcal{A}, \mathcal{P})$, being a (bounded, linear) integral operator, i.e.,

$$Tf(\mathbf{x}_k) = \int f(\mathbf{x}_k, \mathbf{x}_r) f(\mathbf{x}_r) d\mathcal{P}(\mathbf{x}_r),$$

is self-adjoint, Hilbert–Schmidt. With λ's and ν's introduced just before the assumptions, let (ν_s, f_s) forms its orthonormal eigendecomposition, i.e., $h(\mathbf{x}_k, \mathbf{x}_r) = \sum_{s=0}^{\infty} \lambda_s f_s(\mathbf{x}_k) f_s(\mathbf{x}_r)$, where $\sum_s \nu^2 < \infty$ and $f_0 = 1$ correspond to $\lambda_0 = 0$. For details, see e.g., van der Vaart (1998) and Koroljuk and Borovskich (1994). By the Hilbert–Schmidt Theorem (Reed and Simon 1980, p. 203), the convergence of the kernel to its basis is in $\mathcal{L}_2$, i.e.,

$$E\left(h(\mathbf{x}_k, \mathbf{x}_r) - \sum_{s=0}^{p} \lambda_s f_s(\mathbf{x}_k) f_s(\mathbf{x}_r)\right)^2 = \sum_{s=p+1}^{\infty} \nu_s^2 \to 0.$$

A general theorem on the limit of a degenerate U-statistics under this setup is given in van der Vaart (1998, Theorem 12.10, p. 169) or Lee (1990, Theorem 1, p. 90). The limit holds for $n^{c/2}U_{n,c}$ with variance $c!E[h_c^2(\cdot)]$, where $U_{n,c}$ is a U-statistic with (projected) kernel $h_c(\cdot)$ and c is the least value for which $h_c(\cdot)$ is non-degenerate (see Sect. A.1).

Thus, in the present context with $m = 2$, $c = 2$, nU_n has a finite limit with variance approximating $2\xi_2 = 2\,\mathrm{tr}(\boldsymbol{\Sigma}^2)/p^2 = 2\sum_{s=1}^{p}\nu_s^2$. Specifically, for first-order degeneracy, the limit is $[m(m-1)/2]\sum_{s=1}^{\infty}\nu_s(z_s^2-1)$, where z_s are independent $N(0,1)$ variables; see Koroljuk and Borovskich (1994, Ch.) and Shao (2003, Ch. 3). With $m = 2$, we thus have, for $n, p \to \infty$,

$$nU_n \xrightarrow{\mathcal{D}} \sum_{s=1}^{\infty}\nu_s(z_s^2-1), \tag{17}$$

with $z_s^2 \sim \chi_1^2$ iid, where the limiting mean is 0 and variance is $2\sum_{s=1}^{\infty}\nu_s^2$ which approximates $2\,\mathrm{tr}(\boldsymbol{\Sigma}^2)/p^2$. Combined with the limit of nQ_1/p by Slutsky's theorem, we have

$$T - 1 \xrightarrow{\mathcal{D}} \sum_{s=1}^{\infty}\nu_s(z_s^2-1)\Big/\sum_{s=1}^{\infty}\nu_s. \tag{18}$$

Now write $\omega_s = \nu_s/\sum_s\nu_s$ such that $\sum_s\omega_s = 1$ and $\max\omega_s^2 \to 0$. Also let $Y_s = (z_s^2-1)/\sqrt{2}$ so that $E(Y_s) = 0$, $\mathrm{Var}(Y_s) = 1$. Then, the normal limit follows by the Hájek–Šidák Lemma (Lemma 21).

B.2 Proof of Theorem 13

With Q_1 composed of two independent components, the probability convergence of nQ_1/p follows exactly as in one-sample case, so that

$$nQ_1/p \xrightarrow{\mathcal{P}} \nu_0,$$

where $\nu_0 = \sum_{i=1}^{2}\rho_i\nu_{i0} = \sum_{i=1}^{2}\sum_{s=1}^{\infty}\rho_i\nu_{is}$, as $n_i, p \to \infty$. Now Q_0, which we first write as $Q_0 = \mathbf{a}'\mathbf{U}_N$, where $\mathbf{a} = (1\ 1\ -2)'$ and $\mathbf{U}_N = (U_{n_1}\ U_{n_2}\ U_{n_1n_2})'$, so that the limit of Q_0 follows from that of $\mathbf{U}_N$. Obviously $E(Q_0) = \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2$ where, from "Appendix A.2",

$$\mathrm{Cov}(\mathbf{U}_N) = \frac{1}{p^2}\begin{pmatrix} \frac{2\,\mathrm{tr}(\boldsymbol{\Sigma}_1^2)}{n_1(n_1-1)} + \frac{4\boldsymbol{\mu}_1'\boldsymbol{\Sigma}_1\boldsymbol{\mu}_1}{n_1-1} & 0 & \frac{2\boldsymbol{\mu}_2'\boldsymbol{\Sigma}_1\boldsymbol{\mu}_1}{n_1} \\ 0 & \frac{2\,\mathrm{tr}(\boldsymbol{\Sigma}_2^2)}{n_2(n_2-1)} + \frac{4\boldsymbol{\mu}_2'\boldsymbol{\Sigma}_2\boldsymbol{\mu}_2}{n_2-1} & \frac{2\boldsymbol{\mu}_1'\boldsymbol{\Sigma}_2\boldsymbol{\mu}_2}{n_2} \\ \frac{2\boldsymbol{\mu}_2'\boldsymbol{\Sigma}_1\boldsymbol{\mu}_1}{n_1} & \frac{2\boldsymbol{\mu}_1'\boldsymbol{\Sigma}_2\boldsymbol{\mu}_2}{n_2} & \frac{\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)}{n_1n_2} + \frac{\boldsymbol{\mu}_1'\boldsymbol{\Sigma}_2\boldsymbol{\mu}_1}{n_2} + \frac{\boldsymbol{\mu}_2'\boldsymbol{\Sigma}_1\boldsymbol{\mu}_2}{n_1} \end{pmatrix}$$

so that $\mathrm{Var}(Q_0) = \mathbf{a}'\,\mathrm{Cov}(\mathbf{U}_N)\mathbf{a}$ results into

$$\mathrm{Var}(Q_0) = \frac{2}{p^2}\left(\frac{\mathrm{tr}(\boldsymbol{\Sigma}_1^2)}{n_1(n_1-1)} + \frac{\mathrm{tr}(\boldsymbol{\Sigma}_2^2)}{n_2(n_2-1)} + \frac{2\,\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)}{n_1n_2}\right)$$

$$+\frac{4}{p^2}\left(\frac{\boldsymbol{\mu}_1'\boldsymbol{\Sigma}_1\boldsymbol{\mu}_1}{n_1-1}+\frac{\boldsymbol{\mu}_2'\boldsymbol{\Sigma}_2\boldsymbol{\mu}_2}{n_2-1}+\frac{\boldsymbol{\mu}_1'\boldsymbol{\Sigma}_2\boldsymbol{\mu}_1}{n_2}+\frac{\boldsymbol{\mu}_2'\boldsymbol{\Sigma}_1\boldsymbol{\mu}_2}{n_2}\right.$$
$$\left.-\frac{2\boldsymbol{\mu}_2'\boldsymbol{\Sigma}_1\boldsymbol{\mu}_1}{n_1}-\frac{2\boldsymbol{\mu}_1'\boldsymbol{\Sigma}_2\boldsymbol{\mu}_2}{n_2}\right)$$
$$=[2\,\mathrm{tr}(\boldsymbol{\Sigma}_0^2)/p^2+4(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_0(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)/p^2][1+o_P(1)]. \quad (19)$$

Note that, as $n_i, p \to \infty$, terms involving $\boldsymbol{\mu}'\boldsymbol{\Sigma}\boldsymbol{\mu}$ are finite under the assumptions, making $\mathrm{Cov}(n\mathbf{U}_N)$, hence $\mathrm{Var}(nQ_0)$, uniformly bounded, implying in turn that nQ_0 might have a finite limit. Let $\widehat{Q}_0 = \sum_{i=1}^2 \widehat{U}_{n_i} - 2\widehat{U}_{n_1n_2}$ be the projection of $\widetilde{Q}_0 = \sum_{i=1}^2 \widetilde{U}_{n_i} - 2\widetilde{U}_{n_1n_2}$, where $\widetilde{U}_{n_i} = U_{n_i} - E(U_{n_i})$, with its kernel $\widetilde{h}(\mathbf{x}_{ik}, \mathbf{x}_{ir}) = \mathbf{X}_{ik}'\mathbf{X}_{ir}/p - \|\boldsymbol{\mu}\|^2/p$, and similarly $\widetilde{U}_{n_1n_2}$ with its kernel $\widetilde{h}(\mathbf{x}_{1k}, \mathbf{x}_{2l})$. Further, $\widehat{U}_{n_i}$ are as defined for one-sample case, whereas

$$\widehat{U}_{n_1n_2} = \frac{m_1}{n_1}\sum_{k=1}^{n_1}\widetilde{h}_{10}(\mathbf{x}_{1k}) + \frac{m_2}{n_2}\sum_{l=1}^{n_2}\widetilde{h}_{01}(\mathbf{x}_{2l})$$

where $\widetilde{h}_{10}(\mathbf{x}_{1k}) = E[\widetilde{h}(\mathbf{x}_{1k}, \mathbf{x}_{2l})|\mathbf{x}_{1k}]$, similarly $\widetilde{h}_{01}(\mathbf{x}_{2l})$; see Sect. A.1 for notations. Thus

$$\begin{aligned}
\widehat{Q}_0 &= \frac{2}{n_1p}\sum_{k=1}^{n_1}\boldsymbol{\mu}_1'(\mathbf{X}_{1k}-\boldsymbol{\mu}_1) + \frac{2}{n_2p}\sum_{l=1}^{n_2}\boldsymbol{\mu}_2'(\mathbf{X}_{2l}-\boldsymbol{\mu}_2)\\
&\quad -2\left[\frac{1}{n_1p}\sum_{k=1}^{n_1}\boldsymbol{\mu}_2'(\mathbf{X}_{1k}-\boldsymbol{\mu}_1) + \frac{1}{n_2p}\sum_{l=1}^{n_2}\boldsymbol{\mu}_1'(\mathbf{X}_{2l}-\boldsymbol{\mu}_2)\right]\\
&= \frac{2}{n_1p}\sum_{k=1}^{n_1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'(\mathbf{X}_{1k}-\boldsymbol{\mu}_1) - \frac{2}{n_2p}\sum_{l=1}^{n_2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'(\mathbf{X}_{2l}-\boldsymbol{\mu}_2)\\
&= 2(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\left[\frac{1}{n_1p}\sum_{k=1}^{n_1}(\mathbf{X}_{1k}-\boldsymbol{\mu}_1) - \frac{1}{n_2p}\sum_{l=1}^{n_2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\right] \quad (20)
\end{aligned}$$

is the projection of $\widetilde{Q}_0$ with $E(\widehat{Q}_0) = 0$, $\mathrm{Var}(\widehat{Q}_0) = 4(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_0(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)/p^2$. The term within brackets in Eq. (20) is the sum of two independent components, as a direct extension of one-sample case. By the same procedure then, it follows that $\mathrm{Cov}(\widetilde{Q}_0, \widehat{Q}_0) = \mathrm{Var}(\widehat{Q}_0)$ so that $\mathrm{Var}(n\widehat{Q}_0)/\mathrm{Var}(n\widetilde{Q}_0) \to 1$. Under the assumptions, $n\widetilde{Q}_0 = n\widehat{Q}_0 + o_P(1)$ with the limit of $n\widehat{Q}_0$ following by the central limit theorem, leading to the limit of $n\widetilde{Q}_0$, hence of T_2, by Slutsky theorem.

Now consider H_0 whence the projection $\widehat{Q} = 0$, making the component U-statistics, hence Q_0, degenerate and leaving the normal limit above invalid under the null. To simplify the matters, assume without loss of generality, that $\boldsymbol{\mu}_i = \boldsymbol{\mu} = \mathbf{0}$, $i = 1, 2$. Then

$$\mathrm{Cov}(\mathbf{U}_N) = \mathrm{diag}\left(2\,\mathrm{tr}(\boldsymbol{\Sigma}_1^2)/n_1(n_1-1),\quad 2\,\mathrm{tr}(\boldsymbol{\Sigma}_2^2)/n_2(n_2-1),\quad \mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)/n_1n_2\right)/p^2$$

and $\text{Var}(Q_0) = 2\,\text{tr}(\mathbf{\Sigma}_0^2)/p^2$. This, again, is a direct extension of one-sample case under H_0, so that we can similarly proceed to obtain the limit, except that here we need to deal with a three dimensional vector instead of a scalar. Then, the limits of nU_{n_i}, $i = 1, 2$, follow from (17) as (see also Ahmad 2014)

$$nU_{n_i} \xrightarrow{\mathcal{D}} \sum_{s=1}^{\infty} \rho_i \nu_{is}(z_{is}^2 - 1), \tag{21}$$

as $n_i, p \to \infty$, where $n = n_1 + n_2$. Writing $n/n_1n_2 = [\sqrt{n/n_1}\sqrt{n/n_2}][1/\sqrt{n_1n_2}]$, the corresponding limit for $nU_{n_1n_2}$ is given as Koroljuk and Borovskich (1994, Ch. 4)

$$nU_{n_1n_2} \xrightarrow{\mathcal{D}} \sum_{s=1}^{\infty} \sqrt{\rho_1\rho_2}\nu_{1s}\nu_{2s}z_{1s}z_{2s}, \tag{22}$$

as $n_i, p \to \infty$, where z_{is} are iid $N(0, 1)$ variables in both limits and z_{1s}, z_{2s} are also independent of each other. To combine the three limits, define $w_{is}^2 = \rho_i\nu_{is}^2/\sum_s \rho_i\nu_{is}^2$, $i = 1, 2$ such that $\lim_{p\to\infty} \max_s w_{is}^2 = 0$. Then, a multivariate extension of Lemma 21 gives the normal limit

$$\widetilde{\mathbf{U}}_N \xrightarrow{\mathcal{D}} \mathcal{N}_3(\mathbf{0}, \mathbf{I}),$$

where $\widetilde{\mathbf{U}}_N = (U_{n_1}/\sqrt{\text{Var}(U_{n_1})},\ U_{n_2}/\sqrt{\text{Var}(U_{n_21})},\ U_{n_1n_2}/\sqrt{\text{Var}(U_{n_1n_2})})'$ is the standardized form of $\mathbf{U}_N$ with each component having mean zero. Finally, under Assumption 2 and by Slutsky theorem, with covariance matrix diagonal, the limit easily extends for $nQ_0/[nQ_1/p]$ and hence for T_2 as a linear combination of three components.

References

Ahmad, M.R.: A U-statistic approach for a high-dimensional two-sample mean testing problem under non-normality and Behrens–Fisher setting. Ann. Inst. Stat. Math. **66**, 33–61 (2014)

Ahmad, M.R.: On testing sphericity and identity of a covariance matrix with large dimensions. Math. Methods Stat. **25**, 121–132 (2016)

Ahmad, M.R.: Location-invariant multi-sample U-tests for covariance matrices with large dimension. Scand. J. Stat. **44**, 500–523 (2017a)

Ahmad, M.R.: Location-invariant tests of homogeneity of large dimensional covariance matrices. J. Stat. Theory Pract. **11**, 731–745 (2017b)

Anderson, N.H., Hall, P., Titterington, D.M.: Two-sample test statistics for measuring discrepencies between two multivariate probability density functions using kernel based density estimates. J. Multivar. Anal. **50**, 41–54 (1994)

Aoshima, M., Yata, K.: Two-stage procedures for high-dimensional data. Seq. Anal. **30**, 356–399 (2011)

Aoshima, M., Yata, K.: Asymptotic normality for inference on multisample high-dimensional mean vectors under mild conditions. Methodol. Comput. Appl. Probab. **17**, 419–439 (2015)

Armstrong, S.A., Staunton, J.E., Silverman, L.B., et al.: MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nat. Genet. **30**, 41–47 (2002)

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

Beerstecher, E., Sutton, H.E., Berry, H.K., et al.: Biochemical individuality. V. Exlorations with respect to the metabolic pattersn of compulsive drinkers. Arch. Biochem. **29**, 27–40 (1950)

Box, G.E.P.: Some theorems on quadratic forms applied in the study of analysis of variance problems, I: effect of inequality of variance in the one-way classification. Ann. Math. Stat. **25**, 290–302 (1954a)

Box, G.E.P.: Some theorems on quadratic forms applied in the study of analysis of variance problems, II: effect of inequality of variance and of correaltion between errors in the two-way classification. Ann. Math. Stat. **25**, 484–498 (1954b)

Chen, S.X., Qin, Y.-L.: A two-sample test for high-dimensional data with applications to gene-set testing. Ann. Stat. **38**(2), 808–835 (2010)

Dempster, A.P.: A high dimensional two sample significance test. Ann. Math. Stat. **29**, 995–1010 (1958)

Dempster, A.P.: A significance test for the separation of two highly multivariate small samples. Biometrics **16**, 41–50 (1960)

Dempster, A.P.: Elements of Continuous Multivariate Analysis. Addison-Wesley, Reading (1968)

Duchesne, P., Francq, C.: Multivariate hypothesis testing using generalized and 2-inverses with applications. Statistics **49**, 475–496 (2015)

Ellingsen, K.E., Gray, J.S.: Spatial patterns of benthic diversity: is there a latitudinal gradient along the Norwegian continental shelf? J. Anim. Ecol. **71**, 373–389 (2002)

Feng, L., Zou, C., Wang, Z.: Multivariate sign-based high-dimensional test for the two-sample location problem. J. Am. Stat. Assoc. **111**, 721–735 (2016)

Fujikoshi, Y.: Multivariate analysis for the case when the dimension is large compared to the sampel size. J. Korean Stat. Soc. **33**, 1–24 (2004)

Fujikoshi, Y., Ulyanov, V.V., Shimizu, R.: Multivariate statistics: high-dimensional and large-sample approximations. Wiley, New York (2010)

Geissser, S., Greenhouse, W.: An extension of Box's results on the use of F distribution in multivariate analysis. Ann. Math. Stat. **29**, 885–891 (1958)

Greenhouse, W., Geissser, S.: On methods in the analysis of profile data. Psychometrika **24**, 95–112 (1959)

Hájek, J., Šidák, Z., Sen, P.K.: Theory of Rank Tests. Academic Press, Cambridge (1999)

Hu, J., Bai, Z.: A review of 20 years of naive tests of significance for high-dimensional mean vectors and covariance matrices. Sci. China Math. **55**, 1–19 (2015)

Hu, J., Bai, Z., Wang, C., Wang, W.: On testing equality of high dimensional mean vectors with unequal covariance matrices. Ann. Inst. Stat. Math. **69**, 365–387 (2017)

Huynh, H., Feldt, L.S.: Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. J. Am. Stat. Assoc. **65**, 1582–1589 (1970)

Huynh, H., Feldt, L.S.: Estimation of the Box correction for the degrees of freedom from sample data in randomized block and split-plot designs. J. Educ. Stat. **1**, 69–82 (1976)

Jiang, J.: Large Sample Techniques for Statistics. Springer, New York (2010)

Katayama, S., Kano, Y.: A new test on high-dimensional mean vectors without any assumption on population covariance matrix. Commun. Stat. Theory Methods **43**, 5290–5304 (2014)

Koroljuk, V.S., Borovskich, Y.V.: Theory of U-Statistics. Kluwer, Dordrecht (1994)

Läuter, J.: Two new multivariate tests, in particular for a high dimension. Acta et Comment. Univ. Tartu. Math. **8**, 179–186 (2004)

Läuter, J., Glimm, E., Kropf, S.: Multivariate tests based on left-spherically distributed linear scores. Ann. Stat. **26**, 1972–1988 (1998). Corrections: **27**, 1441

Lee, A.J.: U-Statistics: Theory and Practice. CRC, Boca Raton (1990)

Lehmann, E.L.: Elements of Large-Sample Theory. Springer, New York (1999)

Reed, M., Simon, B.: Methods of Modern Mathematical Physics: Functional analysis, vol. I. Academic Press, Cambridge (1980)

Schott, J.R.: Some high-dimensional tests for a one-way MANOVA. J. Multivar. Ann. **98**, 1825–1839 (2007)

Serfling, R.J.: Approximation theorems of Mathematical Statistics. Wiley, Weinheim (1980)

Shao, J.: Mathematical Statistics, 2nd edn. Springer, New York (2003)

van der Vaart, A.W.: Asymptotic Statistics. Cambridge University Press, Cambridge (1998)

Wang, L., Peng, B., Li, R.: A high-dimensional nonparametric multivariate test or mean vector. J. Am. Stat. Assoc. **110**, 1658–1669 (2015)