Received: Added at production | Revised: Added at production | Accepted: Added at production

This is the Pre-Published Version.

DOI: xxx/xxxx

ARTICLE TYPE

Standardized Dempster's non-exact test for high dimensional mean vectors

Hongyan Fang[1] | Yuanyuan Chen[2] | Ling Chen[2] | Wenzhi Yang[2] | Binyan Jiang*[3]

[1]Center for Pure Mathematics, School of Mathematical Sciences, Anhui University, Hefei, Anhui, China
[2]School of Mathematical Sciences, Anhui University, Hefei, Anhui, China
[3]Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China

Correspondence
*Binyan Jiang, Department of Applied Mathematics, The Hong Kong Polytechnic University. Email: by.jiang@polyu.edu.hk

Summary

Although Hotelling's T^2 test has been a widely used test for hypothesis testing problems on the mean vectors, it is not well defined when the data dimension is larger than the sample size. Dempster's non-exact test, as a remedy for the Hotelling's T^2 test, is known to be more powerful than the Hotelling's T^2 test and is well defined even when the dimension is much larger than the sample size. However, Dempster's non-exact test will lose power when the variances of the covariates are different. In this paper, we propose a Standardized Dempster's non-exact test for the classical mean testing problem. The proposed test is more powerful for data with heteroscedastic features, and is applicable to the high dimensional case. An approximate distribution of the test statistic has been established, and to better control the type I error rate when the sample size is small, we further constructed a Monte Carlo version of the proposed standardized Dempster's non-exact test. Various simulation studies and a real data application were conducted with comparison to other popular tests. The numerical results showed that while the type I error rates were well controlled, the testing power of our proposed test was generally higher than those of other tests.

KEYWORDS:
Dempster's non-exact test, Hotelling T^2 test, hypothesis testing, multivariate normal

1 | INTRODUCTION

Comparing the mean vector with a constant vector under the Gaussian assumption arises naturally in a wide range of applications such as, genomics, medical imaging, risk management, and signal processing. The most classical testing method is probably the Hotelling's T^2 test (henceforth HT test) which is well-defined when the vector dimension p is smaller than the sample size n (Hotelling 1992). However, with the rapid development of modern data technology, such as computing and gene sequencing technology, high-dimensional data where the dimension p could be much larger than the sample size n are frequently encountered in many applications. Under the $p > n$ setting, the HT test is no longer applicable, owing to the fact that the sample covariance matrix is no longer invertible. Even when $p < n - 1$, the power of the HT test can be adversely affected if the sample covariance matrix is nearly singular (Bai & Saranadasa 1996; Pan & Zhou 2011; Wang, Peng, & Li 2015).

As a remedy for the ill-conditioned sample covariance matrix under high dimensionality, Dempster proposed a non-exact significance test (henceforth DT test) for the one and two sample mean-comparison problem (Dempster 1958 1960), where the testing statistic is approximately following a F distribution. Bai and Saranadasa (1996) (henceforth BS test) proposed to replace the irreversible sample covariance matrix by the identity matrix under the assumption that $p/n \to \gamma > 0$. Chen and Qin (2010) (henceforth CQ test) modified the BS test and introduced a test statistic by removing the cross-product terms. Pan and Zhou (2011) used the linear spectral statistics to obtain the central limit theorem (CLT) of Hotelling's T^2 test when $p/n \to \gamma \in (0, 1)$. M. S. Srivastava and Du (2008) (henceforth SD test) and M. S. Srivastava (2009) (henceforth S test) used the information from the diagonal elements of the sample covariance matrix to construct their tests when $p \geq n$, and established some related

CLTs. Park and Ayyala (2013) proposed a new scalar transform invariant test and derived the asymptotic null distribution and power under weaker assumptions than those for the S test (M. S. Srivastava 2009). Further related studies on the mean vector testing problem can also be found in Fan and Fan (2008), Thulin (2014), Tony Cai, Liu, and Xia (2014), Liu, Liu, Zheng, and Shi (2017), Wang et al. (2015), Li, Hu, Bai, Yin, and Zou (2017), R. Srivastava, Li, and Ruppert (2016) and Z. Hu, Tong, and Genton (2019), among others.

Despite there are many different tests, Dempster's test has been one of the most useful methods in practice. On one hand, Dempster's test not only serves as a replacement for the Hotelling's T^2 to test the hypothesis when the number of sample is smaller than the dimension, but also is more powerful than the Hotelling's T^2 when the dimension p is moderately large such that Hotelling's T^2 is well defined (J. Hu & Bai 2016). On the other hand, Dempster's test was found to be more powerful than most tests, including the BS test, the CQ test, and the S test (M. S. Srivastava 2007) when the variance of the features are the same. However, if the features are heteroscedastic, i.e., having different variances, Dempster's test would lose power dramatically; see for example Remark 1 in Section 2.1 for more discussions. To deal with the power loss under heteroscedasticity, in this paper, we propose a "standardized Dempster's non-exact test", which generalizes the classical Dempster's non-exact test via a standardization procedure. The standardized Dempster's non-exact test inherits the advantages of the Dempster's non-exact test, and can be applied to the high dimensional circumstance with $p = o(\exp\{cn\})$ for some positive constant c. Although the original Dempster's test is defined for a general covariance matrix where the diagonal elements (i.e., the variances) are not required to be the same, we have found in our numerical studies that the additional standardization step could significantly increase the testing power under heteroscedasticity. We have showed that the proposed test statistic is approximately following an F distribution, and to alleviate the instability of the type I error in the case where the sample size is small, we have further constructed a Monte Carlo type of the standardized Dempster's non-exact test.

In what follows, we first review the Dempster's non-exact test in Section 2.1. The proposed standardized Dempster's non-exact test and related theoretical properties are presented in Section 2.2. To better capture the null distribution and control the type I error rate when the sample size is relatively small, we further propose a Monte Carlo version of our test in Section 2.3. Numerical studies with comparison to other popular tests, including the BS test, the S test, and the DT test are provided in Section 3. Conclusions are provided in Section 4, and all technical proofs are neglected to the Appendix.

1.1 | Methods

1.2 | Retrospection of the Dempster's non-exact test

Let $x_i = (x_{i1}, \cdots, x_{ip})'$, $i = 1, \cdots, n$ be independent and identically distributed (i.i.d.) p-dimensional normal random samples with mean $\mu = (\mu_1, \ldots, \mu_p)'$ and covariance matrix $\Sigma = (\sigma_{ij})_{p\times p}$. Here both μ and Σ are unknown, and Σ is positive definite. Denote the sample data matrix as $X_{n\times p} = (x_1, x_2, \cdots, x_n)'$, i.e., each row of X stands for a sample and each column stands for a variable. Without loss of generality, we consider the following one-sample hypothesis testing problem

$$H_0 : \mu = \mathbf{0} \quad \leftrightarrow \quad H_1 : \mu \neq \mathbf{0} \tag{1}$$

with unknown μ and Σ. The general testing problem $H_0 : \mu = \mu_0$ with a given μ_0 can be converted to the above problem (1) by rewriting $\mu - \mu_0$ as μ.

Let $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = (\bar{x}_1, \ldots, \bar{x}_p)'$ be the sample mean vector, with $\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$, $1 \leq j \leq p$, being the sample means of each component, and denote the sample covariance matrix as

$$S = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})' = (s_{ij})_{p\times p}.$$

Dempster (1958 1960) proposed a non-exact significance test for the hypothesis testing problem (1) via an orthogonal transformation of the data. Specifically, let A be a $n \times n$ orthogonal matrix such that the first row of A is set to be

$$\sqrt{n}\Big(\frac{1}{n}, \frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}\Big).$$

By transforming X into $X^* = AX$, the first row of X^* would become the grand mean of X: $\sqrt{n}\bar{x}' = \sqrt{n}(\bar{x}_1, \ldots, \bar{x}_p)$. Denote the ith row of X^* to be $X_i^{*\prime}$, $i = 1, \cdots, n$. Under this orthogonal transformation, we have:

$$EX_1^{*\prime} = \sqrt{n}\mu' = \sqrt{n}(\mu_1, \ldots, \mu_p),$$

$$EX_i^{*\prime} = \mathbf{0}, \quad 2 \leq i \leq n,$$

$$Var(X_i^*) = \Sigma, \quad 1 \leq i \leq n.$$

Under the null hypothesis, $X_1^*, X_2^*, \cdots, X_n^*$ have the same mean and covariance matrix. So Dempster proposed a significance test

$$T_D = Q_1/[(Q_2 + \cdots + Q_n)/(n-1)], \tag{2}$$

where $Q_i = X_i^{*'} X_i^*$ is the squared length of $X_i^*, i = 1, \cdots, n$. Denote

$$S_h = n\bar{x}\bar{x}' = n \begin{pmatrix} \bar{x}_1^2 & \bar{x}_1\bar{x}_2 & \cdots & \bar{x}_1\bar{x}_p \\ \bar{x}_2\bar{x}_1 & \bar{x}_2^2 & \cdots & \bar{x}_2\bar{x}_p \\ \vdots & \vdots & \vdots & \vdots \\ \bar{x}_p\bar{x}_1 & \bar{x}_p\bar{x}_2 & \cdots & \bar{x}_p^2 \end{pmatrix},$$

and

$$\begin{aligned} S_e &= (n-1)S = (X - 1_{n\times 1}\bar{x}')'(X - 1_{n\times 1}\bar{x}') \\ &= \begin{pmatrix} x_{11}-\bar{x}_1 & x_{21}-\bar{x}_1 & \cdots & x_{n1}-\bar{x}_1 \\ x_{12}-\bar{x}_2 & x_{22}-\bar{x}_2 & \cdots & x_{n2}-\bar{x}_2 \\ \vdots & \vdots & \vdots & \vdots \\ x_{1p}-\bar{x}_p & x_{2p}-\bar{x}_p & \cdots & x_{np}-\bar{x}_p \end{pmatrix} \begin{pmatrix} x_{11}-\bar{x}_1 & x_{12}-\bar{x}_2 & \cdots & x_{1p}-\bar{x}_p \\ x_{21}-\bar{x}_1 & x_{22}-\bar{x}_2 & \cdots & x_{2p}-\bar{x}_p \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1}-\bar{x}_1 & x_{n2}-\bar{x}_2 & \cdots & x_{np}-\bar{x}_p \end{pmatrix}. \end{aligned}$$

We have

$$\mathrm{tr}(S_h) = n\bar{x}'\bar{x} = X_1^{*'} X_1^* = Q_1,$$

and

$$\mathrm{tr}(S_e) = \sum_{i=1}^{n}\sum_{j=1}^{p}(x_{ij} - \bar{x}_j)^2 = \mathrm{tr}(X'X - n\bar{x}'\bar{x}) = \mathrm{tr}(X^*X^{*'}) - Q_1 = Q_2 + \cdots + Q_p.$$

Consequently, Dempster's test statistic (2) reduces to:

$$T_D = (n-1)\mathrm{tr}(S_h)/\mathrm{tr}(S_e) = \frac{n\bar{x}'\bar{x}}{\mathrm{tr}(S)}.$$

Note that T_D appears as the ratio of the between-class variability and the within-class variability, which relates the test to the classical theory in linear regression. Further interpretation of Dempster's test via a multivariate linear model can be found in Fujikoshi, Himeno, and Wakaki (2004) and M. S. Srivastava and Fujikoshi (2006).

When the variables are normally distributed, and under the null hypothesis $\mu = \mathbf{0}$, Dempster showed that $Q_1, \cdots, Q_n$ are independently distributed as a positive quadratic form, and each Q_i can be well approximated by a χ^2-shaped distribution: $Q_i \sim m\chi_r^2$. As a result, we have, approximately,

$$T_D \sim F_{r,(n-1)r}.$$

Here r and m are generally unknown and needs to be estimated (Dempster 1958). Bai and Saranadasa (1996) pointed out that

$$r = \frac{(\mathrm{tr}(\Sigma))^2}{\mathrm{tr}(\Sigma^2)} \quad \text{and} \quad m = \frac{\mathrm{tr}(\Sigma^2)}{\mathrm{tr}(\Sigma)}. \tag{3}$$

From equation (3), we have $r = \frac{(\mathrm{tr}(\Sigma))^2}{\mathrm{tr}(\Sigma^2)} = p\frac{a_1^2}{a_2}$, where $a_i = \mathrm{tr}(\Sigma^i)/p, i = 1, 2$. By replacing Σ with the sample covariance matrix S, we can obtain the following consistent estimators for a_1, a_2 and r:

$$\hat{a}_1 = \frac{\mathrm{tr}(S)}{p}, \quad \hat{a}_2 = \frac{(n-1)^2}{(n-2)(n+1)}\frac{1}{p}\Big[\mathrm{tr}(S^2) - \frac{(\mathrm{tr}(S))^2}{n-1}\Big], \quad \hat{r} = p\frac{\hat{a}_1^2}{\hat{a}_2}. \tag{4}$$

Subsequently, T_D approximately follows the F-distribution with $\lfloor \hat{r} \rfloor$ and $\lfloor (n-1)\hat{r} \rfloor$ degrees of freedom. Here the symbol $\lfloor \cdot \rfloor$ is the largest integer function.

Remark 1. When $\Sigma = \sigma^2 I$, i.e., a homoscedastic case, Bai and Saranadasa (1996) and M. S. Srivastava (2007) established the asymptotic power for Dempster's test, and found that it is more powerful than other tests. Although Dempster's test does not require the variances of the variables to be the same, it would lose power dramatically under heteroscedasticity. To see this, let's take the following two-dimensional case as an example. Consider two scenarios where the means for both scenarios are the same, while the covariance matrix for the first (homoscedastic) scenario is given as $\mathrm{diag}\{3, 3\}$, and the covariance matrix for the other (heteroscedastic) scenario is $\mathrm{diag}\{1, 5\}$. When the sample size is large enough, the sample mean and the trace of the sample covariance matrix for the homoscedastic scenario will be very close to those in the heteroscedastic scenario, resulting in very close values for the test statistic T_D. However, the value of $\mathrm{tr}(S^2)$ for the homoscedastic case would be around 18, while for the heteroscedastic case it would be around 26. Note that the larger value of $\mathrm{tr}(S^2)$ under heteroscedasticity leads to smaller degrees of freedom in the approximated F distribution. Consequently, even though the means for these two scenarios are the same, it is much harder to reject under the heteroscedastic scenario.

1.3 | The standardized Dempster's non-exact test

To alleviate the power loss issue discussed in Remark 1, in the following, we propose a standardized Dempster's non-exact test. Let $V_{p\times p} = \mathrm{diag}(\sigma_{11}, \cdots, \sigma_{pp})$ be the diagonal matrix of variances, with σ_{ii} being the variance of variable $X_i, i = 1, \cdots, p$. We first look at the ideal case by assuming that V is known. Intuitively, we can obtain equal variances by transforming the data into

$$Y = XV^{-\frac{1}{2}} = (y_1, y_2, \cdots, y_n)'.$$

Here $y_i = V^{-\frac{1}{2}} x_i, i = 1, \ldots, n$ are the corresponding transformed samples. We shall use Σ_Y to denote the covariance matrix of y_i. Clearly the y_i's are still independent of each other. Under this standardization, we have

$$\bar{y} = (\bar{y}_1, \cdots, \bar{y}_p)' = \left(\sigma_{11}^{-\frac{1}{2}} \bar{x}_1, \cdots, \sigma_{pp}^{-\frac{1}{2}} \bar{x}_p\right)' = V^{-\frac{1}{2}} \bar{x}.$$

The covariance matrix of data Y is

$$S_Y = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})(y_i - \bar{y})' = V^{-\frac{1}{2}} S V^{-\frac{1}{2}}.$$

Furthermore, we can obtain

$$n\bar{y}'\bar{y} = n(V^{-\frac{1}{2}}\bar{x})'(V^{-\frac{1}{2}}\bar{x}) = n\bar{x}'V^{-1}\bar{x},$$
$$\mathrm{tr}(S_Y) = \mathrm{tr}(V^{-\frac{1}{2}} S V^{-\frac{1}{2}}) = \mathrm{tr}(SV^{-1}).$$

So we can get the ideal standardized Dempster's test statistic for the mean testing problem (1) as follows

$$T_{iSDT} := \frac{n\bar{y}'\bar{y}}{\mathrm{tr}(S_Y)} = \frac{n\bar{x}'V^{-1}\bar{x}}{\mathrm{tr}(SV^{-1})}. \tag{5}$$

Following the so-called χ^2-approximation arguments for the Dempster's non-exact test in Dempster (1958 1960), we immediately have the following result.

Lemma 1. Suppose we have n independent samples $x_i = (x_{i1}, \cdots, x_{ip})' \sim N(\mathbf{0}, \Sigma), i = 1, \cdots, n,$. Let $V = \mathrm{diag}(\Sigma)$ be the diagonal matrix obtained by setting the off-diagonal elements of Σ to be zero. By transforming $Y = XV^{-\frac{1}{2}}$, we approximately have

$$T_{iSDT} = \frac{n\bar{x}'V^{-1}\bar{x}}{\mathrm{tr}(SV^{-1})} \sim F_{r^*,(n-1)r^*},$$

here $r^* = p\frac{b_1^2}{b_2}$, with $b_i = \mathrm{tr}(\Sigma_Y^i)/p = \mathrm{tr}(V^{-\frac{1}{2}}\Sigma V^{-\frac{1}{2}})^i/p, i = 1, 2$.

Practically, the covariance matrix Σ and diagonal matrix V are both unknown, hence T_{iSDT} and the degree parameters in its approximate distribution are unknown. We thus propose to use the following sample analog of T_{iSDT} as the test statistic:

$$T_{SDT} = \frac{n\bar{x}'D^{-1}\bar{x}}{\mathrm{tr}(SD^{-1})} = \frac{n}{p}\bar{x}'D^{-1}\bar{x}. \tag{6}$$

Here $D = \mathrm{diag}(s_{11}, \cdots, s_{pp})$ is the diagonal matrix of the sample variances, and in the last step we have used the fact that $\mathrm{tr}(SD^{-1}) = \mathrm{tr}(D^{-\frac{1}{2}} S D^{-\frac{1}{2}}) = p$. Theorem 1 below shows that when the sample size is large enough, the sample version T_{SDT} is very close to the ideal standardized Dempster's test statistic T_{iSDT}.

Theorem 1. Let $D = \mathrm{diag}(s_{11}, \cdots, s_{pp})$ be the diagonal matrix of the sample variances. Assume that (i) there exists a constant $C > 1$ such that $C^{-1} \leq \lambda_{\min}(\Sigma) \leq \lambda_{\max}(\Sigma) \leq C$, with $\lambda_{\min}(\Sigma)$ and $\lambda_{\max}(\Sigma)$ being the smallest and largest eigenvalues of Σ respectively; (ii) $\frac{\log p}{n} \to 0$. We then have

$$\frac{n\bar{x}'V^{-1}\bar{x}}{\mathrm{tr}(SV^{-1})} - \frac{n}{p}\bar{x}'D^{-1}\bar{x} = o_P(1), \quad n \to \infty.$$

Theorem 1 implies that $T_{SDT} = T_{iSDT} + o_p(1)$, which, by Slutsky's theorem, indicates that the asymptotic distribution of T_{SDT} is the same as that of T_{iSDT}. The condition $\frac{\log p}{n} \to 0$ here implies that our proposed test (6) is valid for the high dimensional case where $p = o(\exp\{cn\})$ for some constant $c > 0$.

Next we derive the sample estimator for the unknown degree parameter r^* of the approximated null distribution in Lemma 1. Note that $b_1 = \frac{\mathrm{tr}(V^{-\frac{1}{2}}\Sigma V^{-\frac{1}{2}})}{p} = 1$. We only need to estimate b_2 and r^*. Similar to (4), we propose to estimator b_2 by:

$$\hat{b}_2 = \frac{(n-1)^2}{(n-2)(n+1)} \frac{1}{p} [\mathrm{tr}(SD^{-1})^2 - \frac{(\mathrm{tr}(SD^{-1}))^2}{n-1}]$$
$$= \frac{(n-1)^2}{(n-2)(n+1)} \frac{1}{p} \sum_{i=1}^{p} \sum_{j=1}^{p} \frac{s_{ij}s_{ji}}{s_{ii}s_{jj}} - \frac{(n-1)p}{(n-2)(n+1)}.$$

Here in the last step we have used the fact that

$$\mathrm{tr}(D^{-\frac{1}{2}}SD^{-\frac{1}{2}}D^{-\frac{1}{2}}SD^{-\frac{1}{2}}) = \mathrm{tr}(SD^{-1}SD^{-1}) = \sum_{i=1}^{p}\sum_{j=1}^{p}\frac{s_{ij}s_{ji}}{s_{ii}s_{jj}}.$$

Then we can estimate the degree parameter r^* by:

$$\hat{r}^* = p\frac{b_1^2}{\hat{b}_2} = \frac{p}{\hat{b}_2}.$$

From Lemma 1 and Theorem 1, we immediately have the following Theorem:

Theorem 2. Let $x_1, x_2, \cdots, x_n \in R^p$ be i.i.d. samples from $N(\mathbf{0}, \Sigma)$, and let the standardized Dempster's test statistic T_{SDT} be defined as in (6). Under the assumptions of Theorem 1, we have approximately,

$$\frac{n}{p}\bar{x}'D^{-1}\bar{x} \sim F_{(\lfloor \hat{r}^* \rfloor, \lfloor (n-1)\hat{r}^* \rfloor)}.$$

For the mean testing problem (1), Theorem 2 indicates that under the null hypothesis, the standardized test statistic T_{SDT} approximately follows the F distribution with degree parameters $(\lfloor \hat{r}^* \rfloor, \lfloor (n-1)\hat{r}^* \rfloor)$.

1.4 | The Monte Carlo type of standardized Dempster's non-exact test

From Theorem 2 we have, the standardized Dempster's non-exact test statistic T_{SDT} converges to statistic T_{iSDT} in probability when sample size $n \to \infty$. However, when the sample size is small, the difference between these two statistics might result in an inaccurate type I error rate. In fact, as we shall see in our simulation studies, when the sample size is small, the type I error rates will be inflated by the extra variability caused by the estimation error of the test statistic. In this circumstance, we propose a Monte Carlo type of the standardized Dempster's test (SDT_{MC}) which can better control the type I error rates.

Based on the expression of T_{SDT} defined in equation (6) and the approximate distribution $F_{(\lfloor \hat{r}^* \rfloor, \lfloor (n-1)\hat{r}^* \rfloor)}$, we can calculate the original p-value: $q_0 = 1 - F_{(\lfloor \hat{r}^* \rfloor, \lfloor (n-1)\hat{r}^* \rfloor)}(T_{SDT})$. The Monte Carlo progress is implemented as follows. Generate n new samples from distribution $N(\mathbf{0}, S)$, where $\mathbf{0}$ is the null hypothesis mean of the population, and S is the sample covariance of initial data. We treat it as the population covariance of the Monte Carlo regenerate data X^*. Calculate the sample mean, the sample covariance matrix of the new data, and obtain a new p-value. We repeat the above Monte Carlo progress for K times independently and denote the p-values as $q_1, q_2, \cdots, q_K$. The final p-value of the SDT_{MC} test is calculated as $p_{mc} = \#\{q_i \le q_0\}/K$. The number of repetitions should be sufficiently large to better capture of the empirical null distribution of T_{SDT}, and $K = 1000$ is used in our numerical studies.

2 | SIMULATION STUDY

In this section, we use various simulation studies to study the performance of the standardized Dempster' non-exact testing method T_{SDT} with comparison to other popular tests. When sample size is relatively small, we also study the performance of the Monte Carlo type of standardized Dempster's non-exact test $T_{SDT_{MC}}$.

The simulation data are generated from the normal distribution $N(\mu, \Sigma)$, here we consider four different variance matrices:

- IHO case (the Independent homoscedastic case): $\Sigma_1 = I$;
- IHE case (the Independent heteroscedastic case): $\Sigma_2 = \begin{pmatrix} I_{p/2} & 0 \\ 0 & 5I_{p/2} \end{pmatrix}$;
- DHO case (Dependent homoscedastic case): $\Sigma_3 = (\rho_{ij})_{p\times p}, \rho_{ij} = \rho^{|i-j|}$;
- DHE case (Dependent heteroscedastic case): $\Sigma_4 = \Sigma_3 - \Sigma_1 + \Sigma_2$. The off-diagonal element of Σ_4 is $\rho_{ij} = \rho^{|i-j|}$, and the diagonal element of Σ_4 is 1 or 5.

The variables in the IHO case are all independent of each other and have equal variances, it is the simplest case. Although the variables in the IHE case are also independent of each other, they have different variances. In the DHO and DHE cases, the variables are dependent, and in the following simulations we choose $\rho = 0.2$. The numerical results in this section are all based on 1 000 replicates and the significance level α is set to be 0.05.

2.1 | Some existing mean tests

As we have introduced in the Introduction section, there are some other classic testing methods for the mean test problem. Besides of the Dempster's non-exact test (DT), we shall compare our proposed tests with the following tests as well:

- BS test: Bai and Saranadasa (1996) proposed the BS test for testing the hypothesis (1). The statistic of BS test is given as

$$T_{BS} = \frac{n\bar{x}'\bar{x} - \text{tr}(S)}{\sqrt{\frac{2n(n-1)}{(n-2)(n+1)}\left[\text{tr}(S)^2 - \frac{(\text{tr}S)^2}{n-1}\right]}}.$$

  Under the null hypothesis, T_{BS} is asymptotically distributed as $N(0, 1)$ under some conditions.

- SD test: Denote $R = D_S^{-\frac{1}{2}} S D^{-\frac{1}{2}}$ to be the sample correlation matrix of data X. M. S. Srivastava and Du (2008) proposed SD test with the test statistic

$$T_{SD} = \frac{n\bar{x}'D^{-1}\bar{x} - (n-1)p/(n-3)}{\sqrt{2(\text{tr}(R^2) - p^2/(n-1))c_{p,n}}},$$

  where the adjustment coefficient $c_{p,n} = 1 + \frac{\text{tr}R^2}{p^{3/2}}$.

- S test: M. S. Srivastava (2009) proposed the S test with test statistic

$$T_S = \frac{n\bar{x}'D^{-1}\bar{x} - (n-1)p/(n-3)}{\sqrt{2(\text{tr}(R^2) - p^2/(n-1))}}.$$

M. S. Srivastava and Du (2008) considered the normality sample of x_i while M. S. Srivastava (2009) discussed the non-normality of x_i without $c_{m,n}$. Under the null hypothesis, T_{SD} and T_S are asymptotically following the $N(0, 1)$ distribution. The two statistics T_{SD}, T_S, and the standardized Dempster' non-exact test all use the information from the diagonal elements of the singular sample covariance matrix S, while T_{BS} chooses to use the sum of the s_{ii}.

2.2 | Type-I error rate

We first evaluate the type I error rates of the standardized Dempster' non-exact testing method, T_{SDT}, the Dempster' non-exact test T_{DT}, and the Monte Carlo type of standardized Dempster's non-exact test $T_{SDT_{MC}}$ when sample size $n \leq 200$. In this part, we consider three different settings for the sample size n and the dimension p: (A) with $p : n = 0.5 : 1$; (B) with $p : n = 1 : 1$ and (C) with $p : n = 1.5 : 1$.

A large number of simulations were conducted with the sample size changing from 20 to 1, 000. The 95% empirical confidence interval for the type I error rate over the 1, 000 replicates is $(0.0365, 0.0635)$. Figure A1 and Figure A2 present the type I error rates under settings (A)-(C) for the IHO and IHE cases. Similarly, the type I error rates of T_{SDT} and T_{DT} for the DHO and DHE cases are provided in Figures A3 and A4.

From Figures A1 to A4 we can see that, for almost all situations, the Dempster' non-exact test can control the estimated type I error rates well. When the sample size is moderately large ($n \geq 200$), the standardized Dempster' non-exact testing method can control the the estimated type I error rates around the nominal levels. But when the sample size is very small, the estimated type I error rates of T_{SDT} appear to be inflated, while the $T_{SDT_{MC}}$ method can control the type I error rates well. These simulation results suggest that as long as the sample size is not too small, the approximate null distribution in Theorem 2 is working reasonable well and the proposed standardized Dempster's test can be applied to the mean test problem. When sample size is relatively small, we can also use $T_{SDT_{MC}}$ method as a supplement method to better control the type I error rate.

2.3 | Power comparisons for large sample

We consider the numerical performance of the standardized Dempster's non-exact testing method T_{SDT}, and mainly compare it with the T_{DT}, T_S, and T_{BS} tests (introduced in Section 3.1) under various parameter settings. The covariance matrix is set separately for the IHO, IHE, DHO, DHE cases. We consider two types of alternative hypothesis: (A1) we set the means of the first p_1 variables to be μ_0, i.e., $\mu' = (\mu_0, \ldots, \mu_0, 0, \ldots, 0)$; (A2) we set the means of the first p_1 variables to be μ_0, and the means of the last p_1 variables to be $-\mu_0$, i.e., $\mu' = (\mu_0, \ldots, \mu_0, 0, \ldots, 0, -\mu_0, \ldots, -\mu_0)$. Case A1 portrays the phenomenon that the non-zero mean has the same changing direction, while case A2 represents the phenomenon that the non-zero mean has different directions. Case A2 is much common in real life. The parameter p_1, representing the non-zero value number of μ, is set as 10% of the total parameter numbers. We set $n = 500$ and 1000 for this power comparison study. When $n = 500$, the variable number is set to be 100, 500, 1000, and when $n = 1000$, we set $p = 500, 1000, 1500$. The different choices of p are corresponding to three different circumstances: $p < n$, $p = n$ and $p > n$. In each parameter setting, different values of μ_0 are chosen to compare the power. Tables 1-4 provide the power comparison results under different parameter settings for the IHO, IHE, DHO, and DHE cases, separately.

TABLE 1 Power comparison at significance level 0.05 for IHO case.

Case		A1					A2				
n	p	μ_0	T_S	T_{BS}	T_{DT}	T_{SDT}	μ_0	T_S	T_{BS}	T_{DT}	T_{SDT}
500	100	0.06	0.272	0.266	0.336	0.351	0.04	0.249	0.251	0.306	0.312
	100	0.08	0.591	0.585	0.658	0.663	0.06	0.692	0.688	0.750	0.756
	500	0.04	0.257	0.251	0.351	0.383	0.03	0.304	0.305	0.405	0.430
	500	0.06	0.760	0.750	0.824	0.838	0.04	0.666	0.665	0.746	0.768
	1000	0.03	0.174	0.170	0.259	0.286	0.02	0.145	0.144	0.218	0.248
	1000	0.05	0.766	0.765	0.850	0.877	0.03	0.502	0.498	0.611	0.644
1000	500	0.03	0.288	0.291	0.387	0.408	0.02	0.250	0.251	0.338	0.346
	500	0.04	0.684	0.683	0.781	0.791	0.03	0.782	0.783	0.844	0.854
	1000	0.02	0.135	0.133	0.204	0.219	0.02	0.421	0.423	0.511	0.539
	1000	0.03	0.481	0.484	0.585	0.605	0.03	0.966	0.964	0.983	0.982
	1500	0.02	0.200	0.203	0.278	0.294	0.015	0.249	0.253	0.333	0.348
	1500	0.03	0.688	0.682	0.769	0.787	0.02	0.572	0.578	0.680	0.697

TABLE 2 Power comparison at significance level 0.05 for IHE case.

Case		A1					A2				
n	p	μ_0	T_S	T_{BS}	T_{DT}	T_{SDT}	μ_0	T_S	T_{BS}	T_{DT}	T_{SDT}
500	100	0.06	0.272	0.066	0.095	0.351	0.06	0.357	0.139	0.178	0.431
	100	0.08	0.591	0.107	0.147	0.663	0.08	0.712	0.270	0.327	0.787
	500	0.04	0.257	0.069	0.091	0.383	0.04	0.340	0.116	0.165	0.466
	500	0.06	0.760	0.127	0.187	0.838	0.05	0.622	0.201	0.279	0.712
	1000	0.04	0.418	0.054	0.101	0.555	0.03	0.237	0.078	0.129	0.359
	1000	0.05	0.766	0.103	0.170	0.877	0.04	0.542	0.164	0.239	0.690
1000	500	0.03	0.288	0.066	0.091	0.408	0.03	0.398	0.136	0.181	0.509
	500	0.04	0.684	0.097	0.153	0.791	0.04	0.827	0.278	0.367	0.890
	1000	0.02	0.135	0.054	0.071	0.219	0.02	0.175	0.083	0.112	0.280
	1000	0.03	0.481	0.082	0.126	0.605	0.03	0.629	0.205	0.285	0.751
	1500	0.02	0.200	0.059	0.080	0.294	0.02	0.264	0.093	0.131	0.370
	1500	0.03	0.688	0.103	0.141	0.787	0.03	0.827	0.257	0.364	0.885

From Tables 1 and 3 we can see that, for the homoscedastic cases, T_{SDT} and T_{DT} have almost the same power, and appear to be more powerful than T_S and T_{BS}. The results for case A1 and case A2 are very similar, which suggests that the direction change of the mean will not affect the power of all the four methods.

The situation is much different for the heteroscedastic case. From Table 2 and Table 4 we can see that the power loss of T_{BS} and T_{DT} is serious. For example, from case A1 in Table 2 we can observe that, when $n = 500$, $p = 1000$ and $\mu_0 = 0.05$, the power of T_{SDT} is 0.877, while the power of T_{BS} and T_{DT} are only 0.103 and 0.170. Overall, our proposed test T_{SDT} is the most powerful method among the four tests for all the parameter settings under the heteroscedastic cases. On the other hand, the power of T_S is also much higher than those of T_{DT} and the T_{BS}. This validates the assertion that the T_S has higher power than T_{DT} and T_{BS} under the heteroscedastic cases (Park & Ayyala 2013; M. S. Srivastava & Du 2008).

2.4 | Power comparisons for small sample

We also compare the power performances of the new methods T_{SDT}, $T_{SDT_{MC}}$ with T_{DT}, T_S and T_{BS} tests for small sample settings. Here we choose sample size $n = 50$, dimension $p = 200$. The Monte Carlo resampling times of $T_{SDT_{MC}}$ are set as $K = 1000$. Different nonzero values of

TABLE 3 Power comparison at significance level 0.05 for DHO case.

Case		A1					A2				
n	p	μ_0	T_S	T_{BS}	T_{DT}	T_{SDT}	μ_0	T_S	T_{BS}	T_{DT}	T_{SDT}
500	100	0.06	0.278	0.269	0.333	0.343	0.05	0.440	0.430	0.491	0.499
	100	0.08	0.548	0.545	0.614	0.627	0.06	0.651	0.645	0.706	0.714
	500	0.04	0.250	0.248	0.321	0.345	0.03	0.291	0.291	0.388	0.405
	500	0.06	0.722	0.721	0.794	0.806	0.04	0.642	0.639	0.718	0.735
	1000	0.04	0.400	0.398	0.502	0.533	0.02	0.147	0.146	0.212	0.240
	1000	0.05	0.718	0.706	0.788	0.829	0.03	0.471	0.470	0.573	0.605
1000	500	0.03	0.273	0.273	0.362	0.365	0.02	0.231	0.226	0.306	0.312
	500	0.04	0.641	0.645	0.721	0.729	0.03	0.730	0.724	0.812	0.818
	1000	0.02	0.139	0.137	0.185	0.198	0.02	0.389	0.390	0.493	0.506
	1000	0.03	0.463	0.459	0.558	0.577	0.025	0.712	0.704	0.807	0.823
	1500	0.02	0.187	0.186	0.257	0.273	0.015	0.230	0.227	0.321	0.340
	1500	0.03	0.639	0.636	0.732	0.749	0.02	0.542	0.541	0.645	0.665

TABLE 4 Power comparison at significance level 0.05 for DHE case.

Case		A1					A2				
n	p	μ_0	T_S	T_{BS}	T_{DT}	T_{SDT}	μ_0	T_S	T_{BS}	T_{DT}	T_{SDT}
500	100	0.06	0.279	0.062	0.094	0.333	0.06	0.360	0.126	0.180	0.428
	100	0.08	0.560	0.106	0.135	0.641	0.08	0.682	0.282	0.343	0.751
	500	0.04	0.262	0.071	0.096	0.352	0.04	0.328	0.118	0.176	0.436
	500	0.06	0.754	0.129	0.179	0.840	0.05	0.615	0.211	0.279	0.718
	1000	0.04	0.399	0.067	0.105	0.543	0.03	0.229	0.08	0.125	0.330
	1000	0.05	0.730	0.102	0.166	0.831	0.04	0.528	0.172	0.254	0.659
1000	500	0.03	0.299	0.061	0.090	0.391	0.03	0.401	0.126	0.184	0.493
	500	0.04	0.664	0.095	0.154	0.743	0.04	0.785	0.288	0.389	0.866
	1000	0.02	0.150	0.056	0.073	0.212	0.02	0.177	0.082	0.105	0.256
	1000	0.03	0.462	0.085	0.126	0.594	0.03	0.620	0.185	0.274	0.720
	1500	0.02	0.204	0.062	0.077	0.297	0.02	0.267	0.094	0.128	0.376
	1500	0.03	0.637	0.096	0.139	0.754	0.03	0.784	0.278	0.384	0.863

mean are chosen for comparison under this small sample study. In particular, the circumstance $\mu_0 = 0$ corresponds to the null hypothesis that all components of mean are 0. The simulation replication is 1000 for all parameter settings. The simulation results are listed in Table 5.

From Table 5, we can see that although the proposed method T_{SDT} obtains the highest power in all the considered methods, its type I error rates inflate under this small size setting. The T_{BS} and T_{DT} can control the type I error rates well. $T_{SDT_{MC}}$ has similar type I error rates as those of T_S, and can better control the type-I error than T_{SDT}. On the other hand, $T_{SDT_{MC}}$ obtains more power than T_S, T_{BS}, and the traditional T_{DT} method for all circumstances.

2.5 | Real data analysis

We apply the four testing methods to the monthly precipitation data of Quebec in Canada. The data covers the monthly mean precipitation of Quebec from April 1943 to December 2018, which can be found at http://climate.weather.gc.ca. Considering the data of the whole year, we adopt the data from January 1944 to December 2018. There are in total 900 samples, with 75 samples for each month. The mean monthly precipitation of Quebec for each month found from the web is $(90, 71, 90, 81, 106, 114, 128, 117, 126, 102, 102, 104)'$ from January to December, and we set it as the mean vector μ_0 in the null hypothesis. Then we treat the January to December monthly precipitation vector of each year as the random

TABLE 5 Power comparison at significance level 0.05 for small samples.

Case	μ_0	A1					μ_0	A2				
		T_S	T_{BS}	T_{DT}	T_{SDT}	$T_{SDT_{MC}}$		T_S	T_{BS}	T_{DT}	T_{SDT}	$T_{SDT_{MC}}$
IHO	0	0.075	0.053	0.049	0.117	0.076	0.10	0.206	0.178	0.230	0.386	0.289
	0.15	0.229	0.204	0.266	0.422	0.349	0.15	0.603	0.565	0.650	0.793	0.714
IHE	0	0.075	0.049	0.049	0.117	0.073	0.15	0.289	0.109	0.141	0.502	0.411
	0.15	0.229	0.068	0.092	0.422	0.346	0.20	0.633	0.201	0.260	0.814	0.736
DHO	0	0.07	0.051	0.048	0.122	0.071	0.10	0.210	0.165	0.229	0.393	0.294
	0.15	0.231	0.200	0.263	0.431	0.325	0.15	0.604	0.575	0.641	0.772	0.699
DHE	0	0.068	0.05	0.051	0.115	0.072	0.15	0.286	0.104	0.148	0.491	0.387
	0.15	0.238	0.062	0.099	0.400	0.321	0.2	0.594	0.216	0.288	0.789	0.702

vector $X_i = (X_{i,1}, \cdots, X_{i,12}), i = 1944, \cdots, 2018$. The mean of X_i is the same as μ. We aim to test the mean hypothesis testing problem

$$H_0 : \mu = \mu_0; \quad \leftrightarrow \quad H_1 : \mu \neq \mu_0.$$

The sample size for each month is 75, which can be treated as small sample testing problem, so we analysis it with $T_{SDT_{MC}}$ also. The p-value of T_S, T_{BS}, T_{DT}, T_{SDT} and $T_{SDT_{MC}}$ are listed in Table 6. All of the five testing methods have similar testing results, and all reject the null hypothesis.

TABLE 6 Real data analysis of monthly precipitation data.

Method	T_S	T_{BS}	T_{DT}	T_{SDT}	$T_{SDT_{MC}}$
p-value	2.78e-6	3.68e-7	4.04e-4	6.58e-4	6e-4

3 | CONCLUSION

In this paper, we propose a standardized Dempster' non-exact testing method T_{SDT} for the mean testing problem of populations, and a Monte Carlo type standardized Dempster' non-exact testing method $T_{SDT_{MC}}$ as a supplement for the small sample case. While the proposed testing method has similar performance as the Dempster' non-exact test when the variances of each variant are the same, it can significantly alleviate the dramatic power loss issue of the Dempster' non-exact test under heteroscedasticity. We have found through our numerical study that the proposed test outperforms other main competitors including T_S and T_{BS}. The approximate distribution of our proposed test T_{SDTT} has been established under the assumption that $\frac{\log p}{n} \to 0$, indicating that our method can be applied to high dimensional data where the dimension p can be much larger than the sample size n. When the sample size is too small, the standardized Dempster' non-exact test may be inflated by the finite sample error, and we have proposed a Monte Carlo version of the standardized Dempster' non-exact test to better recover the null distribution of the proposed standardized Dempster' non-exact test.

ACKNOWLEDGMENTS

This work is partially supported by the National Natural Science Foundation of China (NSFC 11801003), and the Natural Science Foundation of Anhui Province (1808085QA17, 2008085MA14).

□

APPENDIX

A PROOF OF THEOREM 1

For a matrix A, let $\|\mathsf{A}\|_2 = \sqrt{\lambda_{\max}(\mathsf{A}'\mathsf{A})}$ be the spectral norm of A. For the sample mean vector $\bar{\mathsf{x}}$, under the null hypothesis that $\mu = \mathbf{0}$ we have

$$E\bar{x}'\bar{x} = \mathrm{tr}E(\bar{x}\bar{x}') = n^{-1}\mathrm{tr}(\Sigma).$$

By the assumption that $\lambda_{\max}(\boldsymbol{\Sigma}) = \mathsf{O}(1)$, we have

$$\mathrm{tr}\Big(\frac{\Sigma}{n}\Big) = O\Big(\frac{p}{n}\Big).$$

Consequently we have $\frac{\mathsf{n}}{\mathsf{p}}\bar{\mathsf{x}}'\bar{\mathsf{x}} = \mathsf{O}_{\mathsf{P}}(1)$.

Note that $\mathsf{D} - \mathsf{V} = \mathrm{diag}(\mathsf{s}_{11} - \sigma_{11}, \mathsf{s}_{22} - \sigma_{22}, \cdots, \mathsf{s}_{\mathsf{pp}} - \sigma_{\mathsf{pp}})$. From Bernstein's inequality (Jiang 2013), we have for any $\eta > 2$, there exists a large enough constant C_0 such that

$$\mathsf{P}\left(|\mathsf{s}_{\mathsf{ii}} - \sigma_{\mathsf{ii}}| > \mathsf{C}_0\sqrt{\frac{\log(\mathsf{np})}{\mathsf{n}}}\right) \leq (\mathsf{np})^{-\eta}. \tag{A1}$$

Consequently we have

$$\|D - V\|_2 = \lambda_{\max}(D - V) = O_P\left(\sqrt{\frac{\log(np)}{n}}\right) = o_p(1).$$

On the other hand, by condition $\lambda_{\max}(\boldsymbol{\Sigma}) = \mathsf{O}(1)$, we have $\lambda_{\max}(\mathsf{V}) = \mathsf{O}(1)$. Therefore, we have

$$\|D\|_2 \leqslant \|V\|_2 + \|D - V\|_2 = O(1) + O_P\left(\sqrt{\frac{\log(np)}{n}}\right) = O_P(1).$$

Similarly, from (A1) and the fact that $\lambda_{\min}(\boldsymbol{\Sigma}) = \mathsf{O}(1)$, we also have

$$\lambda_{\min}(D) = O_P(1).$$

Therefore, we have

$$\begin{aligned}\left|\frac{\mathsf{n}\bar{\mathsf{x}}'\mathsf{D}^{-1}\bar{\mathsf{x}}}{\mathsf{p}} - \frac{\mathsf{n}}{\mathsf{p}}\bar{\mathsf{x}}'\mathsf{V}^{-1}\bar{\mathsf{x}}\right| &= \frac{\mathsf{n}}{\mathsf{p}}|\bar{\mathsf{x}}'(\mathsf{D}^{-1} - \mathsf{V}^{-1})\bar{\mathsf{x}}| \\ &\leqslant \frac{\mathsf{n}}{\mathsf{p}}\bar{\mathsf{x}}'\bar{\mathsf{x}}\|\mathsf{D}^{-1} - \mathsf{V}^{-1}\|_2 \\ &\leqslant \frac{\mathsf{n}}{\mathsf{p}}\bar{\mathsf{x}}'\bar{\mathsf{x}}\|\mathsf{D}^{-1}\|_2\|\mathsf{V}^{-1}\|_2\|\mathsf{V} - \mathsf{D}\|_2 \\ &= \mathsf{o}_{\mathsf{P}}(1).\end{aligned}$$

Equivalently, we have,

$$\frac{n\bar{x}'D^{-1}\bar{x}}{p} - \frac{n}{p}\bar{x}'V^{-1}\bar{x} = o_P(1). \tag{A2}$$

On the other hand, noticing that $\frac{\mathrm{tr}(\mathsf{SD}^{-1})}{\mathsf{p}} = \frac{\mathrm{tr}(\mathsf{D}^{-1/2}\mathsf{SD}^{-1/2})}{\mathsf{p}} = 1$, we have

$$\begin{aligned}\frac{\mathrm{tr}(\mathsf{SV}^{-1})}{\mathsf{p}} &= \frac{\mathrm{tr}(\mathsf{S}(\mathsf{V}^{-1} - \mathsf{D}^{-1}))}{\mathsf{p}} + \frac{\mathrm{tr}(\mathsf{SD}^{-1})}{\mathsf{p}} \\ &\leq \|\mathsf{V}^{-1} - \mathsf{D}^{-1}\|_2\frac{\mathrm{tr}(\mathsf{S})}{\mathsf{p}} + 1 \\ &= \mathsf{o}_{\mathsf{P}}(1).\end{aligned}$$

Therefore, we have

$$\frac{\mathrm{tr}SV^{-1}}{p} \xrightarrow{P} \frac{\mathrm{tr}SD^{-1}}{p} = 1. \tag{A3}$$

Combining equation (A2) and equation (A3), we have

$$\begin{aligned}\frac{\mathsf{n}\bar{\mathsf{x}}'\mathsf{V}^{-1}\bar{\mathsf{x}}}{\mathrm{tr}(\mathsf{SV}^{-1})} - \frac{\mathsf{n}}{\mathsf{p}}\bar{\mathsf{x}}'\mathsf{D}^{-1}\bar{\mathsf{x}} &= \frac{\mathsf{n}\bar{\mathsf{x}}'\mathsf{V}^{-1}\bar{\mathsf{x}}}{\mathsf{p}}\frac{\mathsf{p}}{\mathrm{tr}(\mathsf{SV}^{-1})} - \frac{\mathsf{n}}{\mathsf{p}}\bar{\mathsf{x}}'\mathsf{D}^{-1}\bar{\mathsf{x}} \\ &= \frac{\mathsf{n}\bar{\mathsf{x}}'\mathsf{V}^{-1}\bar{\mathsf{x}}}{\mathsf{p}}(1 + \mathsf{o}_{\mathsf{P}}(1)) - \frac{\mathsf{n}}{\mathsf{p}}\bar{\mathsf{x}}'\mathsf{D}^{-1}\bar{\mathsf{x}} \\ &= \mathsf{o}_{\mathsf{P}}(1). \quad \#\end{aligned}$$

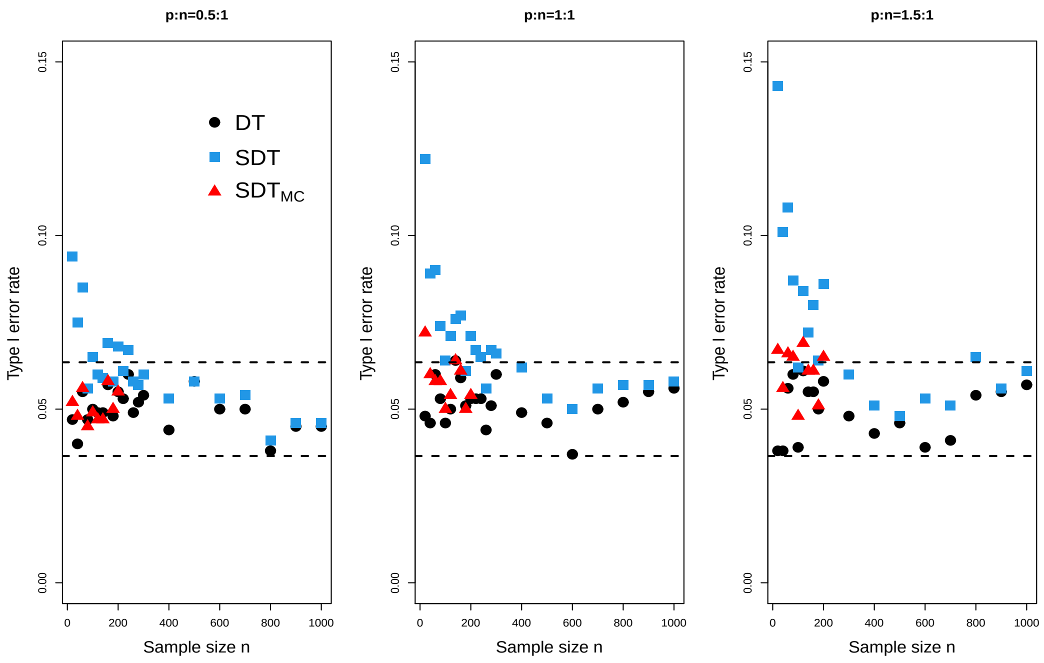

FIGURE A1 Type I error rates of T_{DT}, T_{SDT} and $T_{SDT_{MC}}$ methods for IHO case under significance level $\alpha = 0.05$. The upper and bottom dashed lines represent the lines of type I error equals 0.0365 and 0.0635, respectively.

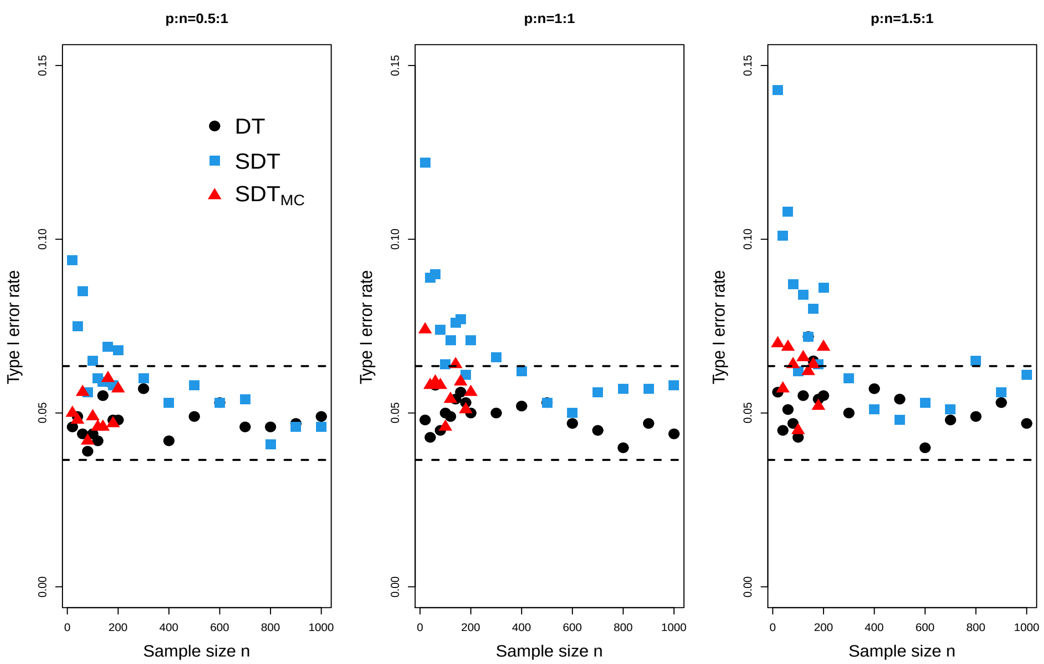

FIGURE A2 Type I error rates of T_{DT}, T_{SDT} and $T_{SDT_{MC}}$ methods for IHE case under significance level $\alpha = 0.05$. The upper and bottom dashed lines represent the lines of type I error equals 0.0365 and 0.0635, respectively.

References

Aitchison, J., & Aitken, C. G. (1976). Multivariate binary discrimination by the kernel method. *Biometrika*, *63*(3), 413–420.

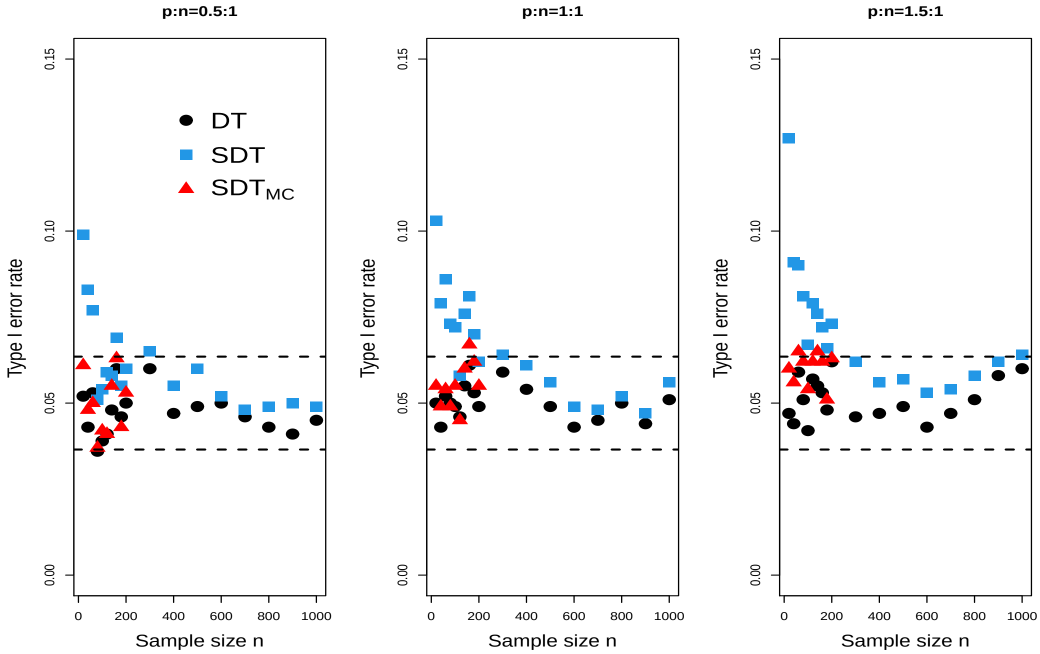

FIGURE A3 Type I error rates of T_{DT}, T_{SDT} and $T_{SDT_{MC}}$ methods for DHO case under significance level $\alpha = 0.05$. The upper and bottom dashed lines represent the lines of type I error equals 0.0365 and 0.0635, respectively.

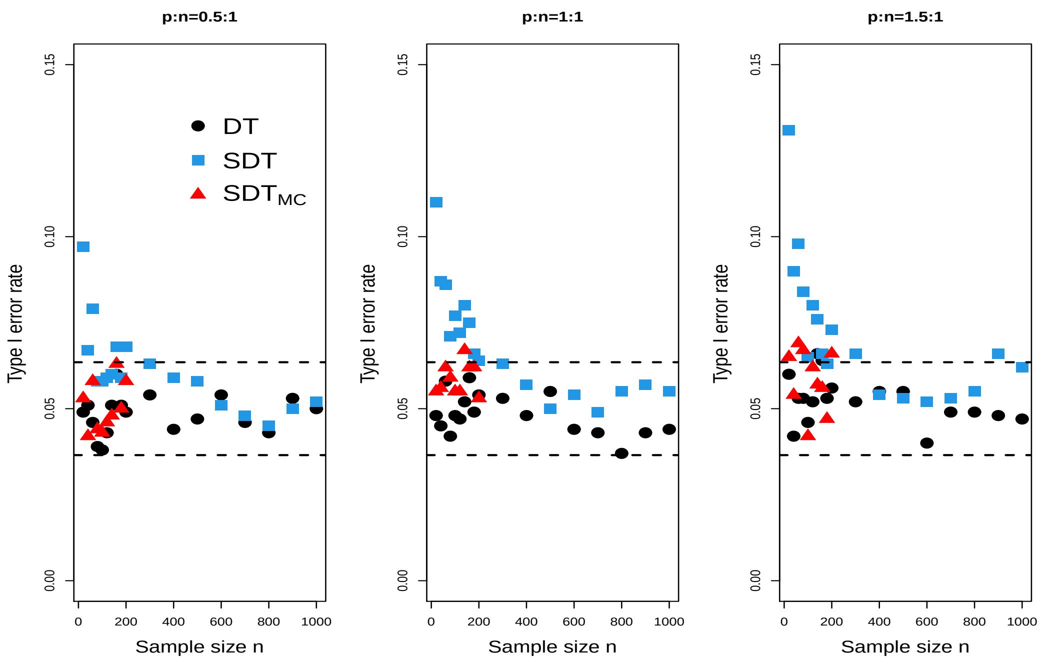

FIGURE A4 Type I error rates of T_{DT}, T_{SDT} and $T_{SDT_{MC}}$ methods for DHE case under significance level $\alpha = 0.05$. The upper and bottom dashed lines represent the lines of type I error equals 0.0365 and 0.0635, respectively.

Anderson, T. (2003). Wiley publications in statistics. *An introduction to multivariate statistical analysis, 3rd ed. Wiley Series in Probability and Statistics. Wiley, New York.*

Asparoukhov, O., & Krzanowski, W. J. (2000). Non-parametric smoothing of the location model in mixed variable discrimination. *Statistics and*

Computing, *10*(4), 289–297.

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

Bickel, P. J., & Levina, E. (2004). Some theory for fisher's linear discriminant function,naive bayes', and some alternatives when there are many more variables than observations. *Bernoulli*, *10*(6), 989–1010.

Bühlmann, P., & Van De Geer, S. (2011). *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media.

Cai, T., & Liu, W. (2011). A direct estimation approach to sparse linear discriminant analysis. *Journal of the American Statistical Association*, *106*(496), 1566–1577.

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

Cochran, W. G., & Hopkins, C. E. (1961). Some classification problems with multivariate qualitative data. *Biometrics*, *17*(1), 10–32.

Daudin, J. (1986). Selection of variables in mixed-variable discriminant analysis. *Biometrics*, 473–481.

De Leon, A., Soo, A., & Williamson, T. (2011). Classification with discrete and continuous variables via general mixed-data models. *Journal of Applied Statistics*, *38*(5), 1021–1032.

Dempster, A. P. (1958). A high dimensional two sample significance test. *The Annals of Mathematical Statistics*, 995–1010.

Dempster, A. P. (1960). A significance test for the separation of two highly multivariate small samples. *Biometrics*, *16*(1), 41–50.

Fan, J., & Fan, Y. (2008). High dimensional classification using features annealed independence rules. *Annals of statistics*, *36*(6), 2605.

Fan, J., Feng, Y., & Tong, X. (2012). A road to classification in high dimensional space: the regularized optimal affine discriminant. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *74*(4), 745–771.

Fan, J., Liu, H., Ning, Y., & Zou, H. (2017). High dimensional semiparametric latent graphical model for mixed data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *79*(2), 405–421.

Fujikoshi, Y., Himeno, T., & 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*(1), 19–26.

Grund, B., & Hall, P. (1993). On the performance of kernel estimators for high-dimensional, sparse binary data. *Journal of Multivariate Analysis*, *44*(2), 321–344.

Gwanoya, T. (2007). Quantitative risk management: Concepts, techniques, tools. by alexander j. mcneil, rüdiger frey & paul embrechts (princeton university press, 2005). *Annals of Actuarial Science*, *2*(1), 187–189.

Hall, P. (1981a). On nonparametric multivariate binary discrimination. *Biometrika*, *68*(1), 287–294.

Hall, P. (1981b). Optimal near neighbour estimator for use in discriminant analysis. *Biometrika*, *68*(2), 572–575.

Hamid, H. (2010). A new approach for classifying large number of mixed variables. In *International conference on computer and applied mathematics* (pp. 156–161).

Hechenbichler, K., & Schliep, K. (2004). Weighted k-nearest-neighbor techniques and ordinal classification.

Hong, S. Y., & Linton, O. (2016). Asymptotic properties of a nadaraya-watson type estimator for regression functions of infinite order. *arXiv preprint arXiv:1604.06380*.

Hotelling, H. (1992). The generalization of student's ratio. In *Breakthroughs in statistics* (pp. 54–65). Springer.

Hu, J., & Bai, Z. (2016). A review of 20 years of naive tests of significance for high-dimensional mean vectors and covariance matrices. *Science China Mathematics*, *59*(12), 2281–2300.

Hu, Z., Tong, T., & Genton, M. G. (2019). Diagonal likelihood ratio test for equality of mean vectors in high-dimensional data. *Biometrics*, *75*(1), 256–267.

Jiang, B. (2013). Covariance selection by thresholding the sample correlation matrix. *Statistics & Probability Letters*, *83*(11), 2492–2498.

Jiang, B., Chen, Z., & Leng, C. (2020). Dynamic linear discriminant analysis in high dimensional space. *Bernoulli*, *26*(2), 1234–1268.

Jiang, B., & Leng, C. (2016). High dimensional discrimination analysis via a semiparametric model. *Statistics & Probability Letters*, *110*, 103–110.

Jiang, B., Wang, X., & Leng, C. (2018). A direct approach for sparse quadratic discriminant analysis. *The Journal of Machine Learning Research*, *19*(1), 1098–1134.

Knoke, J. D. (1982). Discriminant analysis with discrete and continuous variables. *Biometrics*, 191–200.

Kokonendji, C. C., & Ibrahim, M. (2016). Associated kernel discriminant analysis for multivariate mixed data. *Electronic Journal of Applied Statistical Analysis*, *9*(2), 385–399.

Krzanowski, W. (1993). The location model for mixtures of categorical and continuous variables. *Journal of Classification*, *10*(1), 25–49.

Krzanowski, W. (1994). Quadratic location discriminant functions for mixed categorical and continuous data. *Statistics & Probability Letters*, *19*(2), 91–95.

Krzanowski, W. (1995). Selection of variables, and assessment of their performance, in mixed-variable discriminant analysis. *Computational Statistics & Data Analysis*, *19*(4), 419–431.

Krzanowski, W. J. (1975). Discrimination and classification using both binary and continuous variables. *Journal of the American Statistical Association*, *70*(352), 782–790.

Krzanowski, W. J. (1980). Mixtures of continuous and categorical variables in discriminant analysis. *Biometrics*, 493–499.

Li, H., Hu, J., Bai, Z., Yin, Y., & Zou, K. (2017). Test on the linear combinations of mean vectors in high-dimensional data. *Test*, *26*(1), 188–208.

Lin, Z., & Bai, Z. (2011). *Probability inequalities*. Springer Science & Business Media.

Liu, Z., Liu, B., Zheng, S., & Shi, N.-Z. (2017). Simultaneous testing of mean vector and covariance matrix for high-dimensional data. *Journal of Statistical Planning and Inference*, *188*, 82–93.

Mahat, N. I., Krzanowski, W. J., & Hernandez, A. (2007). Variable selection in discriminant analysis based on the location model for mixed variables. *Advances in Data Analysis and Classification*, *1*(2), 105–122.

Mai, Q., Yang, Y., & Zou, H. (2019). Multiclass sparse discriminant analysis. *Statistica Sinica*, *29*, 97–111.

Mai, Q., & Zou, H. (2013). A note on the connection and equivalence of three sparse linear discriminant analysis methods. *Technometrics*, *55*(2), 243–246.

Mai, Q., Zou, H., & Yuan, M. (2012). A direct approach to sparse discriminant analysis in ultra-high dimensions. *Biometrika*, *99*(1), 29–42.

Meier, L., Van De Geer, S., & Bühlmann, P. (2008). The group lasso for logistic regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*(1), 53–71.

Olkin, I., & Tate, R. F. (1961). Multivariate correlation models with mixed discrete and continuous variables. *The Annals of Mathematical Statistics*, *32*(2), 448–465.

Olkin, I., Tate, R. F., et al. (1961). Multivariate correlation models with mixed discrete and continuous variables. *The Annals of Mathematical Statistics*, *32*(2), 448–465.

Pan, G., & Zhou, W. (2011). Central limit theorem for hotelling's t 2 statistic under large dimension. *The Annals of Applied Probability*, 1860–1910.

Park, J., & Ayyala, D. N. (2013). A test for the mean vector in large dimension and small samples. *Journal of Statistical Planning and Inference*, *143*(5), 929–943.

Pereira, B., Chin, S.-F., Rueda, O. M., Vollan, H.-K. M., Provenzano, E., Bardwell, H. A., … others (2016). The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes. *Nature Communications*, *7*(1), 1–16.

Rocha, G. V., Wang, X., & Yu, B. (2009). Asymptotic distribution and sparsistency for ℓ_1-penalized parametric m-estimators with applications to linear svm and logistic regression. *arXiv preprint arXiv:0908.1940*.

Samworth, R. J., et al. (2012). Optimal weighted nearest neighbour classifiers. *The Annals of Statistics*, *40*(5), 2733–2763.

Shao, J., Wang, Y., Deng, X., & Wang, S. (2011). Sparse linear discriminant analysis by thresholding for high dimensional data. *The Annals of Statistics*, *39*(2), 1241–1265.

Srivastava, M. S. (2007). Multivariate theory for analyzing high dimensional data. *Journal of the Japan Statistical Society*, *37*(1), 53–86.

Srivastava, M. S. (2009). A test for the mean vector with fewer observations than the dimension under non-normality. *Journal of Multivariate Analysis*, *100*(3), 518–532.

Srivastava, M. S., & Du, M. (2008). A test for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis*, *99*(3), 386–402.

Srivastava, M. S., & Fujikoshi, Y. (2006). Multivariate analysis of variance with fewer observations than the dimension. *Journal of Multivariate Analysis*, *97*(9), 1927–1940.

Srivastava, R., Li, P., & Ruppert, D. (2016). Raptt: An exact two-sample test in high dimensions using random projections. *Journal of Computational and Graphical Statistics*, *25*(3), 954–970.

Thulin, M. (2014). A high-dimensional two-sample test for the mean using random subspaces. *Computational Statistics & Data Analysis*, *74*, 26–38.

Tony Cai, T., Liu, W., & Xia, Y. (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(2), 349–372.

Wang, L., Peng, B., & Li, R. (2015). A high-dimensional nonparametric multivariate test for mean vector. *Journal of the American Statistical Association*, *110*(512), 1658–1669.

Wei, F., Huang, J., & Li, H. (2011). Variable selection and estimation in high-dimensional varying-coefficient models. *Statistica Sinica*, *21*(4), 1515.

Xu, L., Krzyzak, A., & Suen, C. Y. (1992). Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, *22*(3), 418–435.