The Annals of Statistics
2020, Vol. 48, No. 3, 1304–1328
https://doi.org/10.1214/19-AOS1848

DISTRIBUTION AND CORRELATION-FREE TWO-SAMPLE TEST OF HIGH-DIMENSIONAL MEANS

BY KAIJIE XUE[1] AND FANG YAO[2]

[1]*School of Statistics and Data Science, Nankai University,* kaijie@nankai.edu.cn

[2]*Department of Probability and Statistics, School of Mathematical Sciences, Center for Statistical Science, Peking University,* fyao@math.pku.edu.cn

We propose a two-sample test for high-dimensional means that requires neither distributional nor correlational assumptions, besides some weak conditions on the moments and tail properties of the elements in the random vectors. This two-sample test based on a nontrivial extension of the one-sample central limit theorem (*Ann. Probab.* **45** (2017) 2309–2352) provides a practically useful procedure with rigorous theoretical guarantees on its size and power assessment. In particular, the proposed test is easy to compute and does not require the independently and identically distributed assumption, which is allowed to have different distributions and arbitrary correlation structures. Further desired features include weaker moments and tail conditions than existing methods, allowance for highly unequal sample sizes, consistent power behavior under fairly general alternative, data dimension allowed to be exponentially high under the umbrella of such general conditions. Simulated and real data examples have demonstrated favorable numerical performance over existing methods.

1. Introduction. Two-sample tests of high dimensional means as one of the key issues has attracted a great deal of attention due to its importance in various applications, including [2–5, 10–12, 19, 24–26, 29] and [21], among others. In this article, we tackle this problem with the theoretical advance brought by a high-dimensional two-sample central limit theorem. Based on this, we propose a new type of testing procedure, called distribution and correlation-free (DCF) two-sample mean test, which requires neither distributional nor correlational assumptions and greatly enhances its generality in practice.

We denote two samples by $X^n = \{X_1, \dots, X_n\}$ and $Y^m = \{Y_1, \dots, Y_m\}$ respectively, where X^n is a collection of mutually independent (*not necessarily identically distributed*) random vectors in $\mathbb{R}^p$ with $X_i = (X_{i1}, \dots, X_{ip})'$ and $E(X_i) = \mu^X = (\mu_1^X, \dots, \mu_p^X)'$, $i = 1, \dots, n$, and Y^m is defined in a similar fashion with $E(Y_i) = \mu^Y = (\mu_1^Y, \dots, \mu_p^Y)'$ for all $i = 1, \dots, m$. The normalized sums S_n^X and S_m^Y are denoted by $S_n^X = n^{-1/2}\sum_{i=1}^n X_i = (S_{n1}^X, \dots, S_{np}^X)'$ and $S_m^Y = m^{-1/2}\sum_{i=1}^m Y_i = (S_{m1}^Y, \dots, S_{mp}^Y)'$, respectively. Note that we only assume independent observations, and each sample with a common mean. The hypothesis of interest is

$$H_0 : \mu^X = \mu^Y \quad \text{vs.} \quad H_a : \mu^X \neq \mu^Y,$$

and the proposed two-sample DCF mean test is such that we reject $H_0 : \mu^X = \mu^Y$ at significance level $\alpha \in (0, 1)$, provided that

$$T_n = \|S_n^X - n^{1/2}m^{-1/2}S_m^Y\|_\infty \geq c_B(\alpha),$$

where $T_n = \|S_n^X - n^{1/2}m^{-1/2}S_m^Y\|_\infty$ is the test statistic that only depends on the infinity norm of the sample mean difference, and $c_B(\alpha)$ that plays a central role in this test is a data-dependent critical value defined in (5) of Theorem 3. It is worth mentioning that $c_B(\alpha)$ is easy to

Received October 2018; revised January 2019.
MSC2010 subject classifications. 62H05, 62F05.
Key words and phrases. High-dimensional central limit theorem, Kolmogorov distance, multiplier bootstrap, power function.

computed via a multiplier bootstrap based on a set of independently and identically distributed (i.i.d.) standard normal random variables that are independent of the data, where the explicit calculation is described after (6). Note that the computation of the proposed test is of an order $O\{n(p+N)\}$, more efficient than $O(Nnp)$ that is usually demanded by a general resampling method. In spite of the simple structure of T_n, we shall illustrate its desirable theoretical properties and superior numerical performance in the rest of the article.

We emphasize that our *main contributions* reside on developing a practically useful test that is computationally efficient with rigorous theoretical guarantees given in Theorem 3–5. We begin with deriving nontrivial two-sample extensions of the one-sample central limit theorems and its corresponding bootstrap approximation theorems in high dimensions [9], where we do not require the ratio between sample sizes $n/(n+m)$ to converge but merely reside within an interval (c_1, c_2), $0 < c_1 \le c_2 < 1$, as $n, m \to \infty$. Further, Theorem 3 lays down a foundation for conducting the two-sample DCF mean test uniformly over all $\alpha \in (0,1)$. The power of the proposed test is assessed in Theorem 4 that establishes the asymptotic equivalence between the estimated and true versions. Moreover, the asymptotic power is shown consistent in Theorem 5 under some general alternatives with no sparsity or correlation constraints.

The proposed test sets itself apart from existing methods by allowing for non-i.i.d. random vectors in both samples. The distribution-free feature is in the sense that, under the umbrella of some mild assumptions on the moments and tail properties of the coordinates, there is no other restriction on the distributions of those random vectors. In contrast, existing literature requires the random vectors within sample to be i.i.d. [3–6], and some methods further restrict the coordinates to follow a certain type of distribution, such as Gaussian or sub-Gaussian [26, 29]. This feature sets the proposed test free of making assumptions such as i.i.d. or sub-Gaussianity, which is desirable as distributions of real data are often confounded by numerous factors unknown to researchers. Another key feature is correlation-free in the sense that individual random vectors may have different and arbitrary correlation structures. By contrast, most previous works assume not only a common within-sample correlation matrix, but also some structural conditions, such as those on trace [5], mixing conditions [21] or bounded eigenvalues from below [3]. It is worth noting that our assumptions on the moments and tail properties of the coordinates in random vectors are also weaker than those adopted in literature, for example, [3, 11] and [21] assumed a common fixed upper bound to those moments, [5] and [19] allowed a portion of those moments to go to infinity but paid a price on correlation assumptions.

We also stress that the proposed test possesses consistent power behaviour under fairly general alternative (a mild separation lower bound on $\mu^X - \mu^Y$ in Theorem 5) with neither sparsity nor correlation conditions, while previous works require either sparsity [26] or structural assumption on signal strength [5, 11] or correlation [21], or both [3]. Lastly, we point out that the data dimension p can be exponentially high relative to the sample size under the umbrella of such mild assumptions. This is also favorable compared to previous work, as [3, 5] and [21] allowed such ultrahigh dimensions under nontrivial conditions on either the distribution type (e.g., sub-Gaussian) or the correlation structure (or both) as a tradeoff.

We conclude the Introduction by noting relevant work on one-sample high-dimensional mean tests, such as [14–18, 20, 23, 27, 28] and [1], among others. It is relatively easier to develop a one-sample DCF mean test with similar advantages based on results in [9], thus is not pursued here. The rest of the article is organized as follows. In Section 2, we present the two-sample high-dimensional central limit theorem, and the results on multiplier bootstrap for evaluating the Gaussian approximation. In Section 3, we establish the main result Theorem 3 for conducting the proposed test, and Theorem 4 to approximate its power function, followed by Theorem 5 to analyze its asymptotic power under alternatives. Simulation study is carried out in Section 4 to compare with existing methods, and an application to a real data example is presented in Section 5. We collect the auxiliary lemmas and the proofs of the main results, Theorems 3–5 in the Appendix, and delegate the proofs of Theorems 1–2, Corollary 1 and the auxiliary lemmas to an online Supplementary Material [22] for space economy.

2. Two-sample central limit theorem and multiplier bootstrap in high dimensions. In this section, we first present an intelligible two-sample central limit theorem in high dimensions, which is derived from its more abstract version in Lemma 4 in the Appendix. Then the result on the asymptotic equivalence between the Gaussian approximation adopted in the two-sample central limit theorem and its multiplier bootstrap term is also elaborated, whose abstract version can be referred to Lemma 5.

We first list some notation used throughout the paper. For two vectors $x = (x_1, \dots, x_p) \in \mathbb{R}^p$ and $y = (y_1, \dots, y_p)' \in \mathbb{R}^p$, write $x \le y$ if $x_j \le y_j$ for all $j = 1, \dots, p$. For any $x = (x_1, \dots, x_p)' \in \mathbb{R}^p$ and $a \in \mathbb{R}$, denote $x + a = (x_1 + a, \dots, x_p + a)'$. For any $a, b \in \mathbb{R}$, use the notation $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. For any two sequences of constants a_n and b_n, write $a_n \lesssim b_n$ if $a_n \le C b_n$ up to a universal constant $C > 0$, and $a_n \sim b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$. For any matrix $A = (a_{ij})$, define $\|A\|_\infty = \max_{i,j} |a_{ij}|$. For any function $f : \mathbb{R} \to \mathbb{R}$, write $\|f\|_\infty = \sup_{z \in \mathbb{R}} |f(z)|$. For a smooth function $g : \mathbb{R}^p \to \mathbb{R}$, we adopt indices to represent the partial derivatives for brevity, for example, $\partial_j \partial_k \partial_l g = g_{jkl}$. For any $\alpha > 0$, define the function $\psi_\alpha(x) = \exp(x^\alpha) - 1$ for $x \in [0, \infty)$, then for any random variable X, define

$$\|X\|_{\psi_\alpha} = \inf\{\lambda > 0 : E\{\psi_\alpha(|X|/\lambda)\} \le 1\}, \tag{1}$$

which is an Orlicz norm for $\alpha \in [1, \infty)$ and a quasi-norm for $\alpha \in (0, 1)$.

Denote $F^n = \{F_1, \dots, F_n\}$ as a set of mutually independent random vectors in $\mathbb{R}^p$ such that $F_i = (F_{i1}, \dots, F_{ip})'$ and $F_i \sim N_p(\mu^X, E\{(X_i - \mu^X)(X_i - \mu^X)'\})$ for all $i = 1, \dots, n$, which denotes a Gaussian approximation to X^n. Likewise, define a set of mutually independent random vectors $G^m = \{G_1, \dots, G_m\}$ in $\mathbb{R}^p$ such that $G_i = (G_{i1}, \dots, G_{ip})'$ and $G_i \sim N_p(\mu^Y, E\{(Y_i - \mu^Y)(Y_i - \mu^Y)'\})$ for all $i = 1, \dots, m$ to approximate Y^m. The sets X^n, Y^m, F^n and G^m are assumed to be independent of each other. To this end, denote the normalized sums S_n^X, S_n^F, S_m^Y and S_m^G by $S_n^X = n^{-1/2} \sum_{i=1}^n X_i = (S_{n1}^X, \dots, S_{np}^X)'$, $S_n^F = n^{-1/2} \sum_{i=1}^n F_i = (S_{n1}^F, \dots, S_{np}^F)'$, $S_m^Y = m^{-1/2} \sum_{i=1}^m Y_i = (S_{m1}^Y, \dots, S_{mp}^Y)'$ and $S_m^G = m^{-1/2} \sum_{i=1}^m G_i = (S_{m1}^G, \dots, S_{mp}^G)'$, where S_n^F and S_m^G serve as the Gaussian approximations for S_n^X and S_m^Y, respectively. Lastly, denote a set of independent standard normal random variables $e^{n+m} = \{e_1, \dots, e_{n+m}\}$ that is independent of any of X^n, F^n, Y^m and G^m.

2.1. *Two-sample central limit theorem in high dimensions.* To introduce Theorem 1, a list of useful notation are given as follows. Denote

$$L_n^X = \max_{1 \le j \le p} \sum_{i=1}^n E(|X_{ij} - \mu_j^X|^3)/n, \qquad L_m^Y = \max_{1 \le j \le p} \sum_{i=1}^m E(|Y_{ij} - \mu_j^Y|^3)/m.$$

We denote the key quantity $\rho_{n,m}^{**}$ by

$$\begin{aligned} \rho_{n,m}^{**} = \sup_{A \in \mathcal{A}^{\mathrm{Re}}} \big| & P(S_n^X - n^{1/2}\mu^X + \delta_{n,m} S_m^Y - \delta_{n,m} m^{1/2} \mu^Y \in A) \\ & - P(S_n^F - n^{1/2}\mu^X + \delta_{n,m} S_m^G - \delta_{n,m} m^{1/2} \mu^Y \in A)\big|, \end{aligned} \tag{2}$$

where $P(S_n^X - n^{1/2}\mu^X + \delta_{n,m} S_m^Y - \delta_{n,m} m^{1/2}\mu^Y \in A)$ represents the unknown probability of interest, and $P(S_n^F - n^{1/2}\mu^X + \delta_{n,m} S_m^G - \delta_{n,m} m^{1/2}\mu^Y \in A)$ serves as a Gaussian approximation to this probability of interest, and $\rho_{n,m}^{**}$ measures the error of approximation over all hyperrectangles $A \in \mathcal{A}^{\mathrm{Re}}$. Note that $\mathcal{A}^{\mathrm{Re}}$ is the class of all hyperrectangles in $\mathbb{R}^p$ of the form $\{w \in \mathbb{R}^p : a_j \le w_j \le b_j$ for all $j = 1, \dots, p\}$ with $-\infty \le a_j \le b_j \le \infty$ for all $j = 1, \dots, p$. By assuming more specific conditions, Theorem 1 gives a more explicit bound on $\rho^{**}_{n,m}$ compared to Lemma 4.

THEOREM 1. *For any sequence of constants $\delta_{n,m}$, assume we have the following conditions* (a)–(e):

(a) *There exist universal constants $\delta_1 > \delta_2 > 0$ such that $\delta_2 < |\delta_{n,m}| < \delta_1$.*

(b) *There exists a universal constant $b > 0$ such that*

$$\min_{1\le j\le p} E\{(S^X_{nj} - n^{1/2}\mu^X_j + \delta_{n,m} S^Y_{mj} - \delta_{n,m} m^{1/2}\mu^Y_j)^2\} \ge b.$$

(c) *There exists a sequence of constants $B_{n,m} \ge 1$ such that $L^X_n \le B_{n,m}$ and $L^Y_m \le B_{n,m}$.*

(d) *The sequence of constants $B_{n,m}$ defined in* (c) *also satisfies*

$$\max_{1\le i\le n}\max_{1\le j\le p} E\{\exp(|X_{ij} - \mu^X_j|/B_{n,m})\} \le 2,$$

$$\max_{1\le i\le m}\max_{1\le j\le p} E\{\exp(|Y_{ij} - \mu^Y_j|/B_{n,m})\} \le 2.$$

(e) *There exists a universal constant $c_1 > 0$ such that*

$$(B_{n,m})^2\{\log(pn)\}^7/n \le c_1, \qquad (B_{n,m})^2\{\log(pm)\}^7/m \le c_1.$$

Then we have the following property, where $\rho^{**}_{m,n}$ is defined in (2):

$$\rho^{**}_{n,m} \le K_3([(B_{n,m})^2\{\log(pn)\}^7/n]^{1/6} + [(B_{n,m})^2\{\log(pm)\}^7/m]^{1/6}),$$

for a universal constant $K_3 > 0$.

Conditions (a)–(c) correspond to the moment properties of the coordinates, and (d) concerns the tail properties. It follows from (a) and (b) that the moments *on average* are bounded below away from zero, hence allowing certain proportion of these moments to converge to zero. This is weaker than previous work that requires a uniform lower bound on all moments [3, 11, 21]. Condition (c) implies that the moments *on average* has an upper bound $B_{n,m}$ that can diverge to infinity without restriction on correlation, thus offers more flexibility than those in literature that demands either a fixed upper bound or a certain correlation structure or both. To appreciate this, letting $B_{n,m} \sim n^{1/3}$, one notes that all the variances of the coordinates are allowed to be uniformly as large as $B^{2/3}_{n,m} \sim n^{2/9} \to \infty$ under condition (c), while no restriction on correlation is needed. As a comparison, if we assign a common covariance to two samples, say $\Sigma = (\Sigma_{jk})_{1\le j,k\le p}$ with each $\Sigma_{jk} = n^{2/9}\rho^{1\{j\neq k\}}$ for some constant $\rho \in (0, 1)$, then the trace condition in [5] implies that $p = o(1)$. Compared with a fixed upper bound on the tails of the coordinates [3, 21], condition (d) allows for uniformly diverging tails as long as $B_{n,m} \to \infty$. Condition (e) indicates that the data dimension p can grow exponentially in n, provided that $B_{n,m}$ is of some appropriate order. These conditions as a whole set the basis for the so-called distribution and correlation-free features.

2.2. *Two-sample multiplier bootstrap in high dimensions*. Due to the unknown probability in $\rho^{**}_{n,m}$ (2) denoting the Gaussian approximation, it limits the applicability of the central limit theorem for inference. The idea is to adopt a multiplier bootstrap to approximate its Gaussian approximation, and quantify its approximation error bound. Denote

$$\Sigma^X = n^{-1}\sum_{i=1}^{n} E\{($$

where $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i = (\bar{X}_1, \dots, \bar{X}_p)'$. Analogously, denote Σ^Y, $\hat{\Sigma}^Y$ and $\bar{Y}$. Now we introduce the multiplier bootstrap approximation in this context. Let $e^{n+m} = \{e_1, \dots, e_{n+m}\}$ be a set of i.i.d. standard normal random variables independent of the data, we further denote

$$S_n^{eX} = n^{-1/2}\sum_{i=1}^{n} e_i(X_i - \bar{X}), \qquad S_m^{eY} = m^{-1/2}\sum_{i=1}^{m} e_{i+n}(Y_i - \bar{Y}), \tag{3}$$

and it is obvious that $E_e(S_n^{eX} S_n^{eX\prime}) = \hat{\Sigma}^X$ and $E_e(S_n^{eY} S_n^{eY\prime}) = \hat{\Sigma}^Y$, where $E_e(\cdot)$ means the expectation with respect to e^{n+m} only. Then, for any sequence of constants $\delta_{n,m}$ that depends on both n and m, we denote the quantity of interest $\rho_{n,m}^{MB}$ by

$$\begin{aligned}\rho_{n,m}^{MB} = \sup_{A \in \mathcal{A}^{\mathrm{Re}}} \big| & P_e\big(S_n^{eX} + \delta_{n,m} S_m^{eY} \in A\big) \\ & - P\big(S_n^F - n^{1/2}\mu^X + \delta_{n,m} S_m^G - \delta_{n,m} m^{1/2}\mu^Y \in A\big)\big|,\end{aligned} \tag{4}$$

where $P_e(\cdot)$ means the probability with respect to e^{n+m} only, and $P_e(S_n^{eX} + \delta_{n,m} S_m^{eY} \in A)$ acts as the multiplier bootstrap approximation for the Gaussian approximation $P(S_n^F - n^{1/2}\mu^X + \delta_{n,m} S_m^G - \delta_{n,m} m^{1/2}\mu^Y \in A)$. In particular, $\rho_{n,m}^{MB}$ can be understood as a measure of error between the two approximations over all hyperrectangles $A \in \mathcal{A}^{\mathrm{Re}}$. The following theorem provides a more explicit bound on $\rho_{n,m}^{MB}$ in contrast to its abstract version stated in Lemma 5 in the Appendix.

THEOREM 2. *For any sequence of constants $\delta_{n,m}$, assume we have the following conditions* (a)–(e),

(a) *There exists a universal constant $\delta_1 > 0$ such that $|\delta_{n,m}| < \delta_1$.*

(b) *There exists a universal constant $b > 0$ such that*

$$\min_{1 \le j \le p} E\big\{\big(S_{nj}^X - n^{1/2}\mu_j^X + \delta_{n,m} S_{mj}^Y - \delta_{n,m} m^{1/2}\mu_j^Y\big)^2\big\} \ge b.$$

(c) *There exists a sequence of constants $B_{n,m} \ge 1$ such that*

$$\max_{1 \le j \le p} \sum_{i=1}^{n} E\big\{\big(X_{ij} - \mu_j^X\big)^4\big\}/n \le B_{n,m}^2,$$

$$\max_{1 \le j \le p} \sum_{i=1}^{m} E\big\{\big(Y_{ij} - \mu_j^Y\big)^4\big\}/m \le B_{n,m}^2.$$

(d) *The sequence of constants $B_{n,m}$ defined in* (*c*) *also satisfies*

$$\max_{1 \le i \le n} \max_{1 \le j \le p} E\big\{\exp\big(\big|X_{ij} - \mu_j^X\big|/B_{n,m}\big)\big\} \le 2,$$

$$\max_{1 \le i \le m} \max_{1 \le j \le p} E\big\{\exp\big(\big|Y_{ij} - \mu_j^Y\big|/B_{n,m}\big)\big\} \le 2.$$

(e) *There exists a sequence of constants $\alpha_{n,m} \in (0, e^{-1})$ such that*

$$B_{n,m}^2 \log^5(pn) \log^2(1/\alpha_{n,m})/n \le 1,$$

$$B_{n,m}^2 \log^5(pm) \log^2(1/\alpha_{n,m})/m \le 1.$$

Then there exists a universal constant $c^* > 0$ such that with probability at least $1 - \gamma_{n,m}$ where

$$\begin{aligned}\gamma_{n,m} = {} & (\alpha_{n,m})^{\log(pn)/3} + 3(\alpha_{n,m})^{\log^{1/2}(pn)/c_*} + (\alpha_{n,m})^{\log(pm)/3} \\ & + 3(\alpha_{n,m})^{\log^{1/2}(pm)/c_*} + (\alpha_{n,m})^{\log^3(pn)/6} + 3(\alpha_{n,m})^{\log^3(pn)/c_*} \\ & + (\alpha_{n,m})^{\log^3(pm)/6} + 3(\alpha_{n,m})^{\log^3(pm)/c_*},\end{aligned}$$

we have the following property, where $\rho_{n,m}^{MB}$ is defined in (4),

$$\rho_{n,m}^{MB} \lesssim \{B_{n,m}^2 \log^5(pn)\log^2(1/\alpha_{n,m})/n\}^{1/6} + \{B_{n,m}^2 \log^5(pm)\log^2(1/\alpha_{n,m})/m\}^{1/6}.$$

Conditions (a)–(c) pertain to the moment properties of the coordinates, condition (d) concerns the tail properties and condition (e) characterizes the order of p. These conditions have the desirable features as those in Theorem 1, such as allowing for uniformly diverging moments and tails and so on. Moreover, by combining Theorem 2 with a two-sample Borel–Cantelli lemma (i.e., Lemma 6), where condition (f) is needed for Lemma 6, one can deduce Corollary 1 below, which facilitates the derivation of our main result in Theorem 3.

COROLLARY 1. *For any sequence of constants $\delta_{n,m}$, assume the conditions* (a)–(e) *in Theorem* 2 *hold. Also suppose that the condition* (f) *holds as follows*:

(f) *The sequence of constants $\gamma_{n,m}$ defined in Theorem* 2 *also satisfies*

$$\sum_n \sum_m \gamma_{n,m} < \infty.$$

Then with probability one, *we have the following property*, *where $\rho_{n,m}^{MB}$ is defined in* (4),

$$\rho_{n,m}^{MB} \lesssim \{B_{n,m}^2 \log^5(pn)\log^2(1/\alpha_{n,m})/n\}^{1/6} + \{B_{n,m}^2 \log^5(pm)\log^2(1/\alpha_{n,m})/m\}^{1/6}.$$

3. Two-sample mean test in high dimensions.

In this section, based on the theoretical results from the preceding section, we first establish the main result, Theorem 3, which gives a confidence region for the mean difference $(\mu^X - \mu^Y)$ and, equivalently, the DCF test procedure. We note that the theoretical guarantee is uniform for all $\alpha \in (0, 1)$ with probability one.

THEOREM 3. *Assume we have the following conditions* (a)–(e):

(a) $n/(n+m) \in (c_1, c_2)$, *for some universal constants* $0 < c_1 < c_2 < 1$.

(b) *There exists a universal constant $b > 0$ such that*

$$\min_{1\le j\le p} [E\{(S_{nj}^X - n^{1/2}\mu_j^X)^2\} + E\{(S_{mj}^Y - m^{1/2}\mu_j^Y)^2\}] \ge b.$$

(c) *There exists a sequence of constants $B_{n,m} \ge 1$ such that*

$$\max_{1\le j\le p} \sum_{i=1}^n E(|X_{ij} - \mu_j^X|^{k+2})/n \le B_{n,m}^k,$$

$$\max_{1\le j\le p} \sum_{i=1}^m E(|Y_{ij} - \mu_j^Y|^{k+2})/m \le B_{n,m}^k,$$

for all $k = 1, 2$.

(d) *The sequence of constants $B_{n,m}$ defined in* (*c*) *also satisfies*

$$\max_{1\le i\le n} \max_{1\le j\le p} E\{\exp(|X_{ij} - \mu_j^X|/B_{n,m})\} \le 2,$$

$$\max_{1\le i\le m} \max_{1\le j\le p} E\{\exp(|Y_{ij} - \mu_j^Y|/B_{n,m})\} \le 2.$$

(e) $B_{n,m}^2 \log^7(pn)/n \to 0$ *as* $n \to \infty$.

Then with probability one, the Kolmogorov distance between the distributions of the quantity $\|S_n^X - n^{1/2}m^{-1/2}S_m^Y - n^{1/2}(\mu^X - \mu^Y)\|_\infty$ *and the quantity* $\|S_n^{eX} - n^{1/2}m^{-1/2}S_m^{eY}\|_\infty$ *satisfies*

$$\sup_{t\geq 0}\big|P\big(\|S_n^X - n^{1/2}m^{-1/2}S_m^Y - n^{1/2}(\mu^X - \mu^Y)\|_\infty \leq t\big)$$

$$- P_e\big(\|S_n^{eX} - n^{1/2}m^{-1/2}S_m^{eY}\|$$

It is easy to see that the computation of the DCF test is of the order $O\{n(p+N)\}$, compared with $O(Nnp)$ that is usually demanded by a general resampling method.

According to (6), the power function for the test can be formulated as

$$\text{Power}(\mu^X - \mu^Y) = P\{\|S_n^X - n^{1/2}m^{-1/2}S_m^Y\|_\infty \ge c_B(\alpha) \mid \mu^X - \mu^Y\}. \tag{7}$$

To quantify the power of the DCF test, the expression (7) is not directly applicable since the distribution of $(S_n^X - n^{1/2}m^{-1/2}S_m^Y)$ is unknown. Motivated by Theorem 3, we propose another multiplier bootstrap approximation for $\text{Power}(\mu^X - \mu^Y)$, based on a different set of standard normal random variables $e^{*n+m} = \{e_1^*, \dots, e_{n+m}^*\}$ independent of e^{n+m} that are used to calculate $c_B(\alpha)$,

$$\begin{aligned} &\text{Power}^*(\mu^X - \mu^Y) \\ &\quad = P_{e^*}\{\|S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y} + n^{1/2}(\mu^X - \mu^Y)\|_\infty \ge c_B(\alpha)\}, \end{aligned} \tag{8}$$

where $S_n^{e^*X}$ and $S_m^{e^*Y}$ are as defined in (3) with e^{*n+m} instead of e^{n+m}, and $P_{e^*}(\cdot)$ means the probability with respect to e^{*n+m} only. The following theorem is devoted to establishing the asymptotic equivalence between $\text{Power}(\mu^X - \mu^Y)$ and $\text{Power}^*(\mu^X - \mu^Y)$ under the same conditions as those in Theorem 3.

THEOREM 4. *Assume the conditions* (a)–(e) *in Theorem* 3 *hold*, *then for any* $\mu^X - \mu^Y \in \mathbb{R}^p$, *we have with probability one*,

$$|\text{Power}^*(\mu^X - \mu^Y) - \text{Power}(\mu^X - \mu^Y)| \lesssim \{B_{n,m}^2 \log^7(pn)/n\}^{1/6}.$$

By inspection of the conditions in Theorem 4, it is worth mentioning that neither sparsity nor correlation restriction is required, as opposed to previous works requiring sparsity [3] for instance. To appreciate this point, the asymptotic power under fairly general alternatives specified by condition (f) is analyzed in the theorem below.

THEOREM 5. *Assume the conditions* (a)–(e) *in Theorem* 3 *and that*

(f) $\mathcal{F}_{n,m,p} = \{\mu^X \in \mathbb{R}^p, \mu^Y \in \mathbb{R}^p : \|\mu^X - \mu^Y\|_\infty \ge K_s\{B_{n,m}\log(pn)/n\}^{1/2}\}$, *for a sufficiently large universal constant* $K_s > 0$.

Then for any $\mu^X - \mu^Y \in \mathcal{F}_{n,m,p}$, *we have with probability tending to one*,

$$\text{Power}^*(\mu^X - \mu^Y) \to 1 \quad \textit{as } n \to \infty.$$

The set $\mathcal{F}_{n,m,p}$ in (f) imposes a lower bound on the separation between μ^X and μ^Y, which is comparable to the assumption $\max_i |\delta_i/\sigma_{i,i}^{1/2}| \ge \{2\beta\log(p)/n\}^{1/2}$ in Theorem 2 in [3]. The latter is in fact a special case of condition (f) when the sequence $B_{n,m}$ is constant. It is worth mentioning that the asymptotic power converges to 1 under neither sparsity nor correlation assumptions in the context of our theorem. In contrast, Theorem 2 in [3] requires not only sparse alternatives, but also restrictions on the correlation structure, for example, condition 1 in that theorem such that the eigenvalues of the correlation matrix $\text{diag}(\Sigma)^{-1/2}\Sigma\,\text{diag}(\Sigma)^{-1/2}$ is lower bounded by a positive universal constant. These comparisons reveal that the proposed DCF is powerful for a broader range of alternatives. We conclude this section by noting that the theory for the DCF-type test based on L_2-norm can also be of interest, but is not yet established, which needs further investigation.

4. Simulation studies. In the two-sample test for high-dimensional means, methods that are frequently used and/or recently proposed include those proposed by [5] (abbreviated as CQ, an L_2 norm test), [3] (abbreviated as CL, an L_∞ norm test) and [21] (abbreviated as XL, a test combining L_2 and L_∞ norms) tests. We conduct comprehensive simulation studies to compare our DCF test with these existing methods in terms of size and power under various settings. The two samples $X^n = \{X_i\}_{i=1}^n$ and $Y^m = \{Y_i\}_{i=1}^m$ have sizes (n, m), while the data dimension is chosen to be $p = 1000$. Without loss of generality, we let $\mu^X = 0 \in \mathbb{R}^p$. The structure of $\mu^Y \in \mathbb{R}^p$ is controlled by a signal strength parameter $\delta > 0$ and a sparsity level parameter $\beta \in [0, 1]$. To construct μ^Y, in each scenario, we first generate a sequence of i.i.d. random variables $\theta_k \sim U(-\delta, \delta)$ for $k = 1, \dots, p$ and keep them fixed in the simulation under that scenario. We set $\delta(r) = \{2r \log(p)/(n \vee m)\}^{1/2}$ that gives a proper scale of signal strength [3, 5, 28]. We take $\mu^Y = (\theta_1, \dots, \theta_{\lfloor \beta p \rfloor}, 0'_{p-\lfloor \beta p \rfloor})' \in \mathbb{R}^p$, where $\lfloor a \rfloor$ denotes the nearest integer no more than a, and 0_q is the q-dimensional vector of 0's. Thus the signal becomes sparser for a smaller value of β, with $\beta = 0$ corresponding to the null hypothesis and $\beta = 1$ representing the full dense alternative. The covariance matrices of the random vectors are denoted by $\mathrm{cov}(X_i) = \Sigma^{X_i}$, $\mathrm{cov}(Y_{i'}) = \Sigma^{Y_{i'}}$ for all $i = 1, \dots, n$, $i' = 1, \dots, m$. The nominal significance level is $\alpha = 0.05$, and the DCF test is conducted based on the multiplier bootstrap of size $N = 10^4$.

To have comprehensive comparison, we first consider the following six different settings. The first setting is standard with $(n, m, p) = (200, 300, 1000)$, where the elements in each sample are i.i.d. Gaussian, and the two samples share a common covariance matrix $\Sigma = (\Sigma_{jk})_{1 \le j,k \le p}$. The matrix Σ is specified by a dependence structure such that $\Sigma_{jk} = (1 + |j - k|)^{-1/4}$. Beginning with $\delta = 0.1$, where the implicit chosen value $r = 0.217$ corresponds to quite weak signal according to [3, 28], we calculate the rejection proportions of the four tests based on 1000 Monte Carlo runs over a full range of sparsity levels from $\beta = 0$ (corresponding to null hypothesis) to $\beta = 1$ (corresponding to full dense alternative). Then the signals are gradually strengthened to $\delta = 0.15, 0.2, 0.25, 0.3$. The second setting is similar to the first, except for $\Sigma^{Y_i} = 2\Sigma^{X_{i'}} = 2\Sigma$ for all $i = 1, \dots, n$, $i' = 1, \dots, m$, where Σ is defined in the first setting. These two settings are denoted by "i.i.d. equal (resp., unequal) covariance setting."

In the third setting, the random vectors in each sample have complete different distributions and covariance matrices from one another. The procedure to generate the two samples is as follows. First, a set of parameters $\{\phi_{ij} : i = 1, \dots, m, j = 1, \dots, p\}$ are generated from the uniform distribution $U(1, 2)$ independently, and are kept fixed for all Monte Carlo runs. In a similar fashion, $\{\phi^*_{ij} : i = 1, \dots, m, j = 1, \dots, p\}$ are generated from $U(1, 3)$ independently. Then, for every $i = 1, \dots, n$, we define a $p \times p$ matrix $\Omega_i = (\omega_{ijk})_{1 \le j,k \le p}$ with each $\omega_{ijk} = (\phi_{ij}\phi_{ik})^{1/2}(1 + |j - k|)^{-1/4}$. Likewise, for every $i = 1, \dots, m$, define a $p \times p$ matrix $\Omega^*_i = (\omega^*_{ijk})_{1 \le j,k \le p}$ with each $\omega^*_{ijk} = (\phi^*_{ij}\phi^*_{ik})^{1/2}(1 + |j - k|)^{-1/4}$. Subsequently, we generate a set of i.i.d. random vectors $\breve{X}^n = \{\breve{X}_i\}_{i=1}^n$ with each $\breve{X}_i = (\breve{X}_{i1}, \dots, \breve{X}_{ip})' \in \mathbb{R}^p$, such that $\{\breve{X}_{i1}, \dots, \breve{X}_{i,2p/5}\}$ are i.i.d. standard normal random variables, $\{\breve{X}_{i,2p/5+1}, \dots, \breve{X}_{i,p}\}$ are i.i.d. centered Gamma(16, 1/4) random variables, and they are independent of each other. Accordingly, we construct each X_i by letting $X_i = \mu^X + \Omega_i^{1/2}\breve{X}_i$ for all $i = 1, \dots, n$. It is worth noting that $\Sigma^{X_i} = \Omega_i$ for all $i = 1, \dots, n$, that is, X_i's have different covariance matrices and distributions. The other sample $Y^m = \{Y_i\}_{i=1}^m$ is constructed in the same way with $\Sigma^{Y_i} = \Omega^*_i$ for all $i = 1, \dots, m$. Then we obtained the results for various signal strength levels of δ over a full range of sparsity levels of β, and we denote this setting as "complete relaxed." The fourth setting is analogous to the third, except that we set $(n, m, p) = (100, 400, 1000)$, where two sample sizes deviate substantially from each other. Since this setting is concerned with highly unequal sample sizes, and is therefore denoted as "complete relaxed and highly unequal setting." The fifth setting is similar to the third, except that we replace the standard normal innovations in $\breve{X}_i$ and $\breve{Y}_{i'}$ by independent and heavy-tailed innovations $(5/3)^{-1/2}t(5)$ with mean zero and unit variances, referred to as completely relaxed and heavy-tailed setting. The sixth setting is also analogous to the third, while independent and skewed innovations $8^{-1/2}\{\chi^2(4)-4\}$ with mean zero and unit variances are used, denoted by completely relaxed and skewed setting.

We conduct the four tests and calculate the rejection proportions to assess the empirical power at different signal levels δ and sparsity levels β in each setting as described above, based on 1000 Monte Carlo runs. The numerical results of these six settings are shown in Tables 1–2. For visualization, we depict the empirical power plots of all settings in Figure 1. We also display the multiplier bootstrap approximation based on another independent set of size $N = 10^4$, which agrees well with the empirical size/power of the DCF test and justifies the theoretical assessment in Theorem 4. We see that the empirical sizes of proposed DCF test agree well with the nominal level 0.05 in all six settings. By comparison, the CQ test is not as stable, and the CL and XL tests show underestimation of type I error in all settings.

Regarding power performance under alternatives in these six settings, despite all tests suffering low power for the weak signals $\delta = 0.1$ and $\delta = 0.15$, the DCF test still dominates the other tests at all levels of β. When the signal strength rises to $\delta = 0.2$, the results in Setting I indicate that the DCF test outperforms the other tests, except for the CQ test when $\beta \geq 80\%$ (very dense alternative). Although the power of CQ test increases above that of DCF test at $\beta = 80\%$, the gains are not substantial when both tests have high power. Similar patterns are observed in Settings II, III, V, VI with $\delta = 0.25$ for β ranging between 80% and 83%, and Settings III, IV with $\delta = 0.3$ for β at 80% and 90%, respectively. This phenomenon is visually shown in the power plot in Figure 1. It is also noted the DCF test dominates the CL (L_∞ type) and XL (combined type) uniformly in these settings over all levels of δ and β. To summarize, except for the rapidly increased power of CQ test in very dense alternatives, the DCF test outperforms the other tests over various signal levels of δ in a broad range of sparsity levels β, for alternatives with varied magnitudes and signs. Moreover, the gains are more sustainable in the situations that the data structures get more complex, for example, highly unbalanced sizes, heavy-tailed or skewed distributions.

We further examine alternatives with common/fixed signal upon reviewer's request under the completely relaxed setting, denoted by Setting VII, where we let $\mu^Y = \delta(1, \ldots, 1_{\lfloor \beta p \rfloor}, 0'_{p-\lfloor \beta p \rfloor})'$

TABLE 1
Rejection proportions (%) calculated for four testing methods at different signal strength levels of δ and sparsity levels of β based on 1000 *Monte Carlo runs, where* $\beta = 0$ *corresponds to the null hypothesis* $\beta = 1$ *to the fully dense alternative, and* $(n, m, p) = (200, 300, 1000)$

	Setting I: i.i.d. eu al co																			
	$\delta = 0.1$				$\delta = 0.15$				$\delta = 0.2$				$\delta = 0.25$				$\delta = 0.3$			
Test	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ
$\beta = 0$	4.20	2.40	3.90	5.80	4.30	2.30	2.40	3.60	4.50	2.80	3.70	6.00	4.60	2.70	2.20	3.80	5.00	3.10	3.80	6.10
$\beta = 0.02$	5.00	3.20	2.50	3.40	7.50	4.80	3.70	3.50	15.4	10.5	6.50	3.90	31.7	23.3	14.6	4.40	59.0	47.9	32.6	4.90
$\beta = 0.04$	5.80	3.70	2.80	3.60	10.0	6.20	4.30	3.90	20.6	14.2	8.80	4.70	40.6	30.8	20.0	5.10	72.0	58.9	41.5	5.30
$\beta = 0.2$	9.90	6.50	3.90	4.50	22.7	15.9	9.10	5.30	48.7	37.3	23.7	7.40	84.5	72.4	52.0	11.6	99.3	97.1	87.2	23.4
$\beta = 0.4$	13.9	9.40	5.30	5.20	35.3	25.4	14.4	7.80	68.8	57.1	37.9	16.5	96.8	91.1	72.7	42.5	100	100	97.7	96.9
$\beta = 0.6$	17.8	11.8	6.70	5.60	45.8	33.7	20.3	12.8	82.7	71.8	51.1	39.9	99.6	97.2	86.8	99.1	100	100	100	100
$\beta = 0.8$	22.4	13.8	9.00	8.30	55.5	40.1	24.4	23.1	91.3	81.7	61.5	91.7	100	99.2	95.7	100	100	100	100	100
$\beta = 1$	26.5	17.9	10.9	10.7	64.5	48.1	30.6	39.5	95.0	88.5	70.1	100	100	99.6	100	100	100	100	100	100

	Setting II: i.i.du neu al co																			
	$\delta = 0.1$				$\delta = 0.15$				$\delta = 0.2$				$\delta = 0.25$				$\delta = 0.3$			
Test	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ
$\beta = 0$	4.90	1.80	3.70	6.10	5.20	1.30	2.20	3.80	5.00	1.60	3.60	6.00	4.80	1.20	3.50	6.30	5.00	1.90	3.90	6.20
$\beta = 0.02$	4.70	1.00	2.40	3.80	6.60	1.40	2.70	4.10	10.7	2.60	2.90	4.10	19.1	6.70	4.80	4.40	33.3	14.4	8.80	4.50
$\beta = 0.04$	5.80	1.30	2.50	4.10	7.90	1.80	2.80	4.30	12.5	3.50	3.40	4.50	24.7	9.30	6.00	4.60	42.5	20.3	12.2	5.00
$\beta = 0.2$	8.10	1.90	2.70	4.60	15.0	4.40	3.80	4.90	30.9	11.2	7.20	6.40	57.6	26.5	16.3	8.40	86.8	52.1	33.9	11.8
$\beta = 0.4$	10.6	2.80	3.10	5.70	22.4	7.20	5.70	6.50	47.3	19.6	11.6	10.0	78.7	43.2	26.6	19.1	97.5	74.1	53.2	45.7
$\beta = 0.6$	13.5	3.30	3.80	6.70	29.2	9.60	6.70	8.40	59.0	26.5	17.1	18.7	90.5	56.2	36.7	54.4	99.8	88.1	70.1	99.6
$\beta = 0.8$	16.4	4.60	4.50	7.40	37.4	11.9	8.60	12.6	70.9	32.9	21.4	39.6	95.6	67.0	47.0	1	.	5	7 2	1 2 T

TABLE 1
(*Continued*)

	Setting III: completely relaxed																			
	$\delta = 0.1$				$\delta = 0.15$				$\delta = 0.2$				$\delta = 0.25$				$\delta = 0.3$			
Test	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ
$\beta = 0$	4.70	2.00	3.90	6.30	4.50	1.70	2.30	3.50	4.80	1.90	3.70	6.10	4.60	2.20	2.80	3.90	5.10	2.10	3.80	6.20
$\beta = 0.02$	4.90	2.10	3.20	4.40	6.50	2.70	3.50	5.30	9.40	4.30	4.00	5.60	13.6	7.80	6.20	5.70	24.9	12.9	10.1	5.90
$\beta = 0.04$	5.60	2.40	3.50	4.70	7.60	3.40	4.20	5.40	12.1	6.00	5.00	5.80	19.1	10.8	8.80	6.00	32.8	19.1	13.8	6.50
$\beta = 0.2$	7.50	3.80	4.30	5.80	12.1	6.00	5.60	6.60	23.9	12.5	8.90	7.50	44.2	26.3	16.6	9.30	71.6	50.2	32.1	14.1
$\beta = 0.4$	9.40	3.90	4.50	6.30	18.4	9.00	8.00	7.60	35.8	19.9	12.7	11.7	62.3	40.8	26.4	18.5	89.3	69.9	48.6	31.5
$\beta = 0.6$	11.5	4.90	6.20	6.80	24.0	10.8	8.90	9.50	48.0	28.2	18.2	17.8	76.8	55.3	37.0	35.7	96.5	83.8	64.6	83.1
$\beta = 0.8$	13.6	6.40	6.60	7.00	30.3	13.5	11.7	12.7	57.3	36.4	23.4	28.5	86.7	65.0	45.1	81.2	98.5	91.6	77.4	100
$\beta = 0.83$	14.3	7.10	6.80	7.50	31.0	14.6	11.8	13.1	58.0	37.6	23.9	30.8	87.6	66.1	46.1	88.0	98.9	92.6	79.2	100
$\beta = 1$	16.6	8.50	7.40	8.00	35.0	17.2	13.9	17.3	65.6	42.8	28.3	48.2	90.8	75.7	56.0	99.9	99.2	95.5	95.7	100

TABLE 2

Rejection proportions (%) calculated for four testing methods at different signal strength levels of δ and sparsity levels of β based on 1000 *Monte Carlo runs, where $\beta = 0$ corresponds to the null hypothesis $\beta = 1$ to the fully dense alternative, $(n, m, p) = (100, 400, 1000)$ for Setting IV, and $(n, m, p) = (200, 300, 1000)$ for Settings V and VI*

Setting IV: completely relaxed and highly unequal sample sizes

	$\delta = 0.1$				$\delta = 0.15$				$\delta = 0.2$				$\delta = 0.25$				$\delta = 0.3$			
Test	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ
$\beta = 0$	4.70	0.800	3.90	6.80	4.90	0.900	3.80	6.30	5.20	0.700	3.90	6.10	4.50	0.600	3.50	6.00	4.90	0.500	3.40	6.10
$\beta = 0.02$	5.20	1.10	2.90	4.70	5.90	1.00	3.60	5.60	6.70	1.40	4.60	5.80	8.90	2.40	5.00	5.80	13.2	4.20	6.20	5.90
$\beta = 0.04$	5.40	1.20	3.00	4.80	6.30	1.30	4.50	5.70	7.80	1.90	5.00	6.00	11.2	3.30	5.60	6.10	17.6	5.70	7.10	6.20
$\beta = 0.2$	6.60	1.30	3.30	5.40	9.20	2.20	5.10	5.80	14.9	3.90	5.70	6.20	25.3	8.70	7.00	7.50	42.8	16.5	11.8	8.80
$\beta = 0.4$	7.80	2.00	4.30	5.50	12.4	3.40	5.20	6.10	22.3	6.60	7.10	8.60	38.2	13.0	9.70	10.7	61.3	24.8	17.0	15.8
$\beta = 0.6$	9.10	2.40	4.60	5.80	16.1	3.80	5.50	7.90	29.5	10.0	9.20	10.8	49.9	19.3	14.3	17.6	75.3	33.7	21.9	34.2
$\beta = 0.8$	10.5	2.50	4.70	6.10	19.9	5.20	6.70	9.20	36.9	12.7	10.9	14.5	60.1	24.0	19.3	32.2	84.9	46.6	33.6	78.2
$\beta = 0.9$	11.3	2.80	4.80	6.40	21.9	5.40	7.10	9.90	39.5	13.3	12.6	17.7	64.6	26.6	21.6	43.8	88.0	48.6	35.3	94.0
$\beta = 1$	12.1	2.90	5.30	7.30	23.4	5.90	7.30	11.0	42.0	14.6	12.8	21.7	68.6	29.6	24.5	59.0	90.9	53.1	41.9	99.4

Setting V: completely relaxed and heavy-tailed

	$\delta = 0.1$				$\delta = 0.15$				$\delta = 0.2$				$\delta = 0.25$				$\delta = 0.3$			
Test	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ
$\beta = 0$	4.20	2.20	3.80	6.20	5.20	2.50	3.90	6.10	4.70	1.90	2.90	6.00	4.30	2.00	1.70	3.90	4.50	2.30	2.00	3.70
$\beta = 0.02$	5.50	2.10	3.70	5.40	6.40	2.50	3.90	5.50	9.50	4.40	4.60	6.10	15.3	7.40	6.30	6.10	25.5	15.0	10.3	6.20
$\beta = 0.04$	6.20	2.30	3.80	5.50	7.20	3.60	4.20	6.00	12.6	6.60	5.80	6.20	18.9	9.80	7.00	6.50	33.3	20.7	13.0	7.10
$\beta = 0.2$	7.50	3.60	4.00	5.80	12.4	6.80	6.50	7.30	23.5	13.0	9.60	8.90	45.6	27.6	17.9	11.3	71.7	52.6	33.8	14.1
$\beta = 0.4$	9.50	4.20	4.40	5.90	18.1	9.00	8.30	8.90	35.9	21.3	14.0	12.7	64.4	43.2	26.9	18.5	90.3	73.4	52.0	33.7
$\beta = 0.6$	11.5	5.10	4.50	6.00	23.8	12.6	10.1	11.7	46.7	29.2	19.4	17.8	77.5	55.9	37.4	38.9	97.4	86.5	65.6	88.2
$\beta = 0.8$	13.7	7.30	6.20	8.80	29.4	16.0	12.3	14.1	56.5	36.9	24.9	28.9	87.4	69.1	48.3	81.4	99.2	93.6	80.0	100
$\beta = 0.83$	14.1	7.50	6.30	9.20	30.6	17.3	13.0	15.2	58.1	38.1	26.0	32.0	88.1	70.1	49.5	87.5	99.3	94.1	82.1	100
$\beta = 1$	16.1	8.90	7.40	9.40	34.9	18.9	15.0	17.2	64.5	44.6	30.5	52.2	91.6	75.1	56.6	99.8	99.7	96.5	96.0	100

TABLE 2
(*Continued*)

	Setting VI: completely relaxed and skewed																			
	$\delta = 0.1$				$\delta = 0.15$				$\delta = 0.2$				$\delta = 0.25$				$\delta = 0.3$			
Test	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ	DCF	CL	XL	CQ
$\beta = 0$	4.20	2.10	2.40	3.60	4.90	1.40	2.70	3.80	5.00	1.60	2.50	3.90	4.90	2.40	3.70	5.80	4.70	1.90	2.70	3.90
$\beta = 0.02$	4.80	1.30	2.70	4.40	6.20	1.70	3.10	4.70	7.50	2.70	3.80	4.90	12.9	5.80	5.00	5.00	24.3	11.8	8.30	5.00
$\beta = 0.04$	5.30	1.40	3.00	4.60	7.00	2.30	3.30	4.90	11.3	5.20	4.50	5.10	17.1	8.70	7.00	5.10	32.2	17.3	12.0	5.30
$\beta = 0.2$	7.40	3.00	3.30	4.80	12.8	5.80	5.00	5.80	23.0	12.9	9.20	6.40	42.4	25.6	17.7	8.40	71.3	48.6	32.5	12.4
$\beta = 0.4$	9.40	4.50	4.00	5.10	18.7	9.30	6.80	7.20	37.3	21.9	13.4	10.6	62.9	43.3	28.6	17.3	89.4	70.9	51.8	30.7
$\beta = 0.6$	11.5	5.70	4.50	6.20	24.7	12.3	9.60	9.50	48.1	29.8	18.1	16.5	75.7	55.0	37.6	34.8	95.9	83.7	64.5	86.4
$\beta = 0.8$	14.2	6.30	5.80	6.60	30.5	14.9	10.5	12.5	58.0	37.6	23.4	27.1	86.7	65.4	44.9	80.2	98.7	92.0	77.5	100
$\beta = 0.83$	14.3	7.50	6.	.	.3.	12														

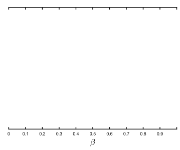
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
β

TABLE 3
Shown are the results of four tests based the original dataset, the bootstrapped samples and the random permutations

	p-values of the four tests based on the dataset			
Test	DCF	CL	XL	CQ
p-value	0.006	0.1708	0.093	0.0955

	Rejection proportions (%) of the four tests over 500 bootstrapped datasets			
Test	DCF	CL	XL	CQ
Rejection proportion	82	65.8	65	58

	Rejection proportions (%) of the four tests over 500 random permutations			
Test	DCF	CL	XL	CQ
Rejection proportion	4.6	1.8	3.4	7.4

500 bootstrapped datasets are given in Table 3, which shows that the highest rejection proportion among the four tests is achieved by DCF at 82%. This is in line with the smallest and significant p-value given by the DCF test based on the dataset itself. We also perform 500 random permutations of the whole dataset (i.e., mixing up two groups that eliminate the group difference) and conduct four tests over each permuted dataset. From Table 3, we see that the rejection proportion of the DCF test (0.046) is close to the nominal level $\alpha = 0.05$, while those of the other tests differ considerably.

APPENDIX

We first present some auxiliary lemmas that are key for deriving the main theorems. To introduce Lemma 1, for any $\beta > 0$ and $y \in \mathbb{R}^p$, we define a function $F_\beta(w)$ as

$$F_\beta(w) = \beta^{-1} \log\left[\sum_{j=1}^{p} \exp\{\beta(w_j - y_j)\}\right], \quad w \in \mathbb{R}^p,$$

which satisfies the property

$$0 \le F_\beta(w) - \max_{1\le j\le p}(w_j - y_j) \le \beta^{-1}\log p,$$

for every $w \in \mathbb{R}^p$ by (1) in [8]. In addition, we let $\varphi_0 : \mathbb{R} \to [0, 1]$ be a real-valued function such that φ_0 is thrice continuously differentiable and $\varphi_0(z) = 1$ for $z \le 0$ and $\varphi_0(z) = 0$ for $z \ge 1$. For any $\phi \ge 1$, define a function $\varphi(z) = \varphi_0(\phi z)$, $z \in \mathbb{R}$. Then, for any $\phi \ge 1$ and $y \in \mathbb{R}^p$, denote $\beta = \phi \log p$ and define a function $\kappa : \mathbb{R}^p \to [0, 1]$ as

$$\kappa(w) = \varphi_0(\phi F_{\phi \log p}(w)) = \varphi(F_\beta(w)), \quad w \in \mathbb{R}^p. \tag{9}$$

Lemma 1 is devoted to characterize the properties of the function κ defined in (9), which can be also referred to Lemmas A.5 and A.6 in [7].

LEMMA 1. *For any $\phi \ge 1$ and $y \in \mathbb{R}^p$, we denote $\beta = \phi \log p$, then the function κ defined in (9) has the following properties, where κ_{jkl} denotes $\partial_j\partial_k\partial_l\kappa$. For any $j, k, l = 1, \ldots, p$, there exists a nonnegative function Q_{jkl} such that*:

(1) $|\kappa_{jkl}(w)| \leq Q_{jkl}(w)$ *for all* $w \in \mathbb{R}^p$,

(2) $\sum_{j=1}^p \sum_{k=1}^p \sum_{l=1}^p Q_{jkl}(w) \lesssim (\phi^3 + \phi^2\beta + \phi\beta^2) \lesssim \phi\beta^2$ *for all* $w \in \mathbb{R}^p$,

(3) $Q_{jkl}(w) \lesssim Q_{jkl}(w + \tilde{w}) \lesssim Q_{jkl}(w)$ *for all* $w \in \mathbb{R}^p$ *and* $\tilde{w} \in \{w^* \in \mathbb{R}^p : \max_{1\leq j\leq p} |w_j^*|\beta \leq 1\}$.

To state Lemma 2, a two-sample extension of Lemma 5.1 in [9], for any sequence of constants $\delta_{n,m}$ that depends on both n and m, we denote the quantity $\rho_{n,m}$ by

$$\begin{aligned}\rho_{n,m} = \sup_{v\in[0,1]} \sup_{y\in\mathbb{R}^p} &\big|P\big\{v^{1/2}\big(S_n^X - n^{1/2}\mu^X + \delta_{n,m}S_m^Y - \delta_{n,m}m^{1/2}\mu^Y\big) \\ &+ (1-v)^{1/2}\big(S_n^F - n^{1/2}\mu^X + \delta_{n,m}S_m^G - \delta_{n,m}m^{1/2}\mu^Y\big) \leq y\big\} \\ &- P\big(S_n^F - n^{1/2}\mu^X + \delta_{n,m}S_m^G - \delta_{n,m}m^{1/2}\mu^Y \leq y\big)\big|.\end{aligned} \tag{10}$$

Lemma 2 provides a bound on $\rho_{n,m}$ under some general conditions.

[redacted] ≥ 1 *and any sequence of constants* $\delta_{n,m}$, *assume the following* [redacted]

(a) *There exists a universal constant* $b > 0$ *such that*

$$\min_{1\leq j\leq p} E\big\{\big(S_{nj}^X - n^{1/2}\mu_j^X + \delta_{n,m}S_{mj}^Y - \delta_{n,m}m^{1/2}\mu_j^Y\big)^2\big\} \geq b.$$

Then we have

$$\rho_{n,m} \lesssim n^{-1/2}\phi_1^2(\log p)2$$

Then we have

$$\begin{aligned}\rho^*_{n,m} \le K^*\big[&n^{-1/2}\phi_1^2(\log p)^2\{\phi_1 L_n^X \rho^*_{n,m} + L_n^X(\log p)^{1/2} + \phi_1 M_n(\phi_1)\}\\ &+ m^{-1/2}\phi_2^2(\log p)^2|\delta_{n,m}|^3\{\phi_2 L_m^Y\rho^*_{n,m} + L_m^Y(\log p)^{1/2} + \phi_2 M^*_m(\phi_2)\}\\ &+ \big(\min\{\phi_1,\phi_2\}\big)^{-1}(\log p)^{1/2}\big],\end{aligned}$$

up to a universal constant $K^* > 0$ that depends only on b, where $\rho^*_{n,m}$ is defined in (11).

Before stating the next lemma, for any $\phi \ge 1$, we denote $M_n(\phi) = M_n^X(\phi) + M_n^F(\phi)$, where $M_n^X(\phi)$ and $M_n^F(\phi)$ are given as follows, respectively,

$$n^{-1}\sum_{i=1}^n E\Big[\max_{1\le j\le p}|X_{ij} - \mu_j^X|^3 1\Big\{\max_{1\le j\le p}|X_{ij} - \mu_j^X| > n^{1/2}/(4\phi\log p)\Big\}\Big],$$

$$n^{-1}\sum_{i=1}^n E\Big[\max_{1\le j\le p}|F_{ij} - \mu_j^F|^3 1\Big\{\max_{1\le j\le p}|F_{ij} - \mu_j^F| > n^{1/2}/(4\phi\log p)\Big\}\Big],$$

similar to those adopted in [9]. Likewise, for any $\phi \ge 1$ and any sequence of constants $\delta_{n,m}$ that depends on both n and m, we denote $M^*_m(\phi) = M^Y_m(\phi) + M^G_m(\phi)$ with $M^Y_m(\phi)$ and $M^G_m(\phi)$ as follows, respectively,

$$m^{-1}\sum_{i=1}^m E\Big[\max_{1\le j\le p}|Y_{ij} - \mu_j^Y|^3 1\Big\{\max_{1\le j\le p}|Y_{ij} - \mu_j^Y| > m^{1/2}/\big(4|\delta_{n,m}|\phi\log p\big)\Big\}\Big],$$

$$m^{-1}\sum_{i=1}^m E\Big[\max_{1\le j\le p}|G_{ij} - \mu_j^G|^3 1\Big\{\max_{1\le j\le p}|G_{ij} - \mu_j^G| > m^{1/2}/\big(4|\delta_{n,m}|\phi\log p\big)\Big\}\Big].$$

Recalling the definition of $\rho^{**}_{n,m}$ in (2), Lemma 4 gives an abstract upper bound on $\rho^{**}_{n,m}$ under mild conditions as follows.

LEMMA 4. *For any sequence of constants $\delta_{n,m}$, assume we have the following conditions* (a)–(b):

(a) *There exists a universal constant $b > 0$ such that*

$$\min_{1\le j\le p} E\{(S^X_{nj} - n^{1/2}\mu^X_j + \delta_{n,m}S^Y_{mj} - \delta_{n,m}m^{1/2}\mu^Y_j)^2\} \ge b.$$

(b) *There exist two sequences of constants $\bar{L}^*_n$ and $\bar{L}^{**}_m$ such that we have $\bar{L}^*_n \ge L^X_n$ and $\bar{L}^{**}_m \ge L^Y_m$, respectively. Moreover, we also have*

$$\phi^*_n = K_1\{(\bar{L}^*_n)^2(\log p)^4/n\}^{-1/6} \ge 2,$$

$$\phi^{**}_m = K_1\{(\bar{L}^{**}_m)^2(\log p)^4|\delta_{n,m}|^6/m\}^{-1/6} \ge 2,$$

for a universal constant $K_1 \in (0, (K^* \vee 2)^{-1}]$, where the positive constant K^* that depends on n as defined in Lemma 3 *in the Appendix*.

Then we have the following property, where $\rho^{**}_{n,m}$ is defined in (2),

$$\begin{aligned}\rho^{**}_{n,m} \le K_2\big[&\{(\bar{L}^*_n)^2(\log p)^7/n\}^{1/6} + \{M_n(\phi^*_n)/\bar{L}^*_n\}\\ &+ \{(\bar{L}^{**}_m)^2(\log p)^7|\delta_{n,m}|^6/m\}^{1/6} + \{M^*_m(\phi^{**}_m)/\bar{L}^{**}_m\}\big],\end{aligned}$$

for a universal constant $K_2 > 0$ that depends only on b.

To introduce Lemma 5, for any sequence of constants $\delta_{n,m}$ that depends on both n and m, denote a useful quantity $\hat{\Delta}_{n,m} = \|\hat{\Sigma}^X - \Sigma^X + \delta_{n,m}^2(\hat{\Sigma}^Y - \Sigma^Y)\|_\infty$. Lemma 5 below gives an abstract accuracy bound on $\rho_{n,m}^{MB}$ defined in (4).

LEMMA 5. *For any sequence of constants $\delta_{n,m}$, assume we have the following condition* (a):

(a) *There exists a universal constant $b > 0$ such that*

$$\min_{1\le j\le p} E\{(S_{nj}^X - n^{1/2}\mu_j^X + \delta_{n,m}S_{mj}^Y - \delta_{n,m}m^{1/2}\mu_j^Y)^2\} \ge b.$$

Then for any sequence of constants $\bar{\Delta}_{n,m} > 0$, on the event $\{\hat{\Delta}_{n,m} \le \bar{\Delta}_{n,m}\}$, we have the following property, where $\rho_{n,m}^{MB}$ is defined in (4),

$$\rho_{n,m}^{MB} \lesssim (\bar{\Delta}_{n,m})^{1/3}(\log p)^{2/3}.$$

Lastly, we present the two-sample Borel–Cantelli lemma in Lemma 6.

LEMMA 6. *Let $\{A_{n,m} : n \ge 1, m \ge 1, (n,m) \in A\}$ be a sequence of events in the sample space Ω, where A is the set of all possible combinations (n,m), which has the form $A = \{(n,m) : n \ge 1, m \in \sigma(n)\}$ where $\sigma(n)$ is a set of positive integers determined by n, possibly the empty set. Assume the following condition* (a):

(a) $\sum_{n=1}^{\infty}\sum_{m\in\sigma(n)} P(A_{n,m}) < \infty$.

Then we have the following property:

$$P\left(\bigcap_{k_1=1}^{\infty}\bigcap_{k_2=1}^{\infty}\bigcup_{n=k_1}^{\infty}\bigcup_{m\in\varrho(k_2)\cap\sigma(n)} A_{n,m}\right) = 0,$$

where $\varrho(k_2) = \{k : k \in \mathbb{N}, k \ge k_2\}$.

Note that if $m \in \sigma(n) = \varnothing$, we just delete the roles of those $A_{n,m}$ and $A_{n,m}^c$ during any operations such as union and intersection, and the same applies to $P(A_{n,m})$ and $P(A_{n,m}^c)$ during summation and deduction.

Before proceeding, we mention that the derivations of Theorems 1–2 essentially follow those of their counterparts in [9], but need more technicalities to employ the aforesaid Lemmas 4–5 to address the challenge arising from unequal sample sizes. The derivation of Corollary 1 is based on Theorem 1 as well as a two-sample Borel–Cantelli lemma (Lemma 6) that first appears in this work as far as we know.

Theorems 3–5 regarding the DCF test are newly developed, while no comparable results are present in literature. Thus we present the proofs of Theorems 3–5 below, while the proofs of Theorems 1–2, Corollary 1 and the auxiliary lemmas are delegated to an online Supplementary Material for space economy.

PROOF OF THEOREM 3. First of all, we define a sequence of constants $\delta_{n,m}$ by

$$\delta_{n,m} = -n^{1/2}m^{-1/2}. \tag{12}$$

Together with condition (a), it can deduced that

$$\delta_2 < |\delta_{n,m}| < \delta_1, \tag{13}$$

ith $\delta_1 = \{c_2/(1-c_2)\}^{1/2} > 0$ and $\delta_2 = \{c_1/(1-c_1)\}^{1/2} >$

PROOF OF THEOREM 4. Gi en an $(\mu^X - \mu^Y)$, e ha e

$$\begin{aligned}
&\text{Po e}^*(\mu^X - \mu^Y) \\
&= P_{e^*}\{\|S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y} + n^{1/2}(\mu^X - \mu^Y)\|_\infty \geq c_B(\alpha)\} \\
&= 1 - P_{e^*}\{\|S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y} + n^{1/2}(\mu^X - \mu^Y)\|_\infty < c_B(\alpha)\} \\
&= 1 - P_{e^*}\{-n^{1/2}(\mu^X - \mu^Y) - c_B(\alpha) < S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y} < \\
&\quad -n^{1/2}(\mu^X - \mu^Y) + c_B(\alpha)\} \\
&= 1 - P_{e^*}\{-n^{1/2}(\mu^X - \mu^Y) - c_B(\alpha) < S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y} < \\
&\quad -n^{1/2}(\mu^X - \mu^Y) + c_B(\alpha)\} \\
&\quad + P\{-n^{1/2}(\mu^X - \mu^Y) - c_B(\alpha) < S_n^X - n^{1/2}m^{-1/2}S_m^Y \\
&\quad - n^{1/2}(\mu^X - \mu^Y) < -n^{1/2}(\mu^X - \mu^Y) + c_B(\alpha)\} \\
&\quad - P\{-n^{1/2}(\mu^X - \mu^Y) - c_B(\alpha) < S_n^X - n^{1/2}m^{-1/2}S_m^Y \\
&\quad - n^{1/2}(\mu^X - \mu^Y) < -n^{1/2}(\mu^X - \mu^Y) + c_B(\alpha)\} \\
&\geq 1 - \sup_{A \in \mathcal{A}^{\mathrm{Re}}} |P(\|S_n^X - n^{1/2}m^{-1/2}S_m^Y \\
&\quad - n^{1/2}(\mu^X - \mu^Y)\|_\infty \in A) - P_{e^*}(\|S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y}\|_\infty \in A)| \\
&\quad - P\{\|S_n^X - n^{1/2}m^{-1/2}S_m^Y\|_\infty < c_B(\alpha)\} \\
&= \text{Po e}(\mu^X - \mu^Y) \\
&\quad - \sup_{A \in \mathcal{A}^{\mathrm{Re}}} |P(\|S_n^X - n^{1/2}m^{-1/2}S_m^Y - n^{1/2}(\mu^X - \mu^Y)\|_\infty \in A) \\
&\quad - P_{e^*}(\|S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y}\|_\infty \in A)|.
\end{aligned} \tag{22}$$

Like ise, gi en an $(\mu^X - \mu^Y)$, e ha e

$$\begin{aligned}
&\text{Po e}(\mu^X - \mu^Y) \\
&= P\{\|S_n^X - n^{1/2}m^{-1/2}S_m^Y\|_\infty \geq c_B(\alpha)\} \\
&= 1 - P\{\|S_n^X - n^{1/2}m^{-1/2}S_m^Y\|_\infty < c_B(\alpha)\} \\
&= 1 - P\{-c_B(\alpha) < S_n^X - n^{1/2}m^{-1/2}S_m^Y < c_B(\alpha)\} \\
&= 1 + P_{e^*}\{-n^{1/2}(\mu^X - \mu^Y) - c_B(\alpha) < S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y} < \\
&\quad -n^{1/2}(\mu^X - \mu^Y) + c_B(\alpha)\} - P\{-n^{1/2}(\mu^X - \mu^Y) - c_B(\alpha) \\
&\quad < S_n^X - n^{1/2}m^{-1/2}S_m^Y - n^{1/2}(\mu^X - \mu^Y) < -n^{1/2}(\mu^X - \mu^Y) + c_B(\alpha)\} \\
&\quad - P_{e^*}\{-n^{1/2}(\mu^X - \mu^Y) - c_B(\alpha) < S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y} \\
&\quad < -n^{1/2}(\mu^X - \mu^Y) + c_B(\alpha)\} \\
&\geq 1 - \sup_{A \in \mathcal{A}^{\mathrm{Re}}} |P(\|S_n^X - n^{1/2}m^{-1/2}S_m^Y - n^{1/2}(\mu^X - \mu^Y)\|_\infty \in A) \\
&\quad - P_{e^*}(\|S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y}\|_\infty \in A)|
\end{aligned} \tag{23}$$

$$
\begin{aligned}
&- P_{e^*}\{\|S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y} + n^{1/2}(\mu^X - \mu^Y)\|_\infty < c_B(\alpha)\} \\
&= \mathrm{Power}^*(\mu^X - \mu^Y) \\
&\quad - \sup_{A \in \mathcal{A}^{\mathrm{Re}}} \big| P(\|S_n^X - n^{1/2}m^{-1/2}S_m^Y - n^{1/2}(\mu^X - \mu^Y)\|_\infty \in A) \\
&\quad - P_{e^*}(\|S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y}\|_\infty \in A)\big|.
\end{aligned}
$$

Putting (22) and (23) together indicates that

$$
\begin{aligned}
&\big|\mathrm{Power}^*(\mu^X - \mu^Y) - \mathrm{Power}(\mu^X - \mu^Y)\big| \\
&\quad \le \sup_{A \in \mathcal{A}^{\mathrm{Re}}} \big| P(\|S_n^X - n^{1/2}m^{-1/2}S_m^Y - n^{1/2}(\mu^X - \mu^Y)\|_\infty \in A) \\
&\quad - P_{e^*}(\|S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y}\|_\infty \in A)\big|.
\end{aligned} \tag{24}
$$

Moreover, by similar argument as in the proof of Theorem 3, one can show that with probability one,

$$
\begin{aligned}
&\sup_{A \in \mathcal{A}^{\mathrm{Re}}} \big| P(\|S_n^X - n^{1/2}m^{-1/2}S_m^Y - n^{1/2}(\mu^X - \mu^Y)\|_\infty \in A) \\
&\quad - P_{e^*}(\|S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y}\|_\infty \in A)\big| \\
&\quad \lesssim \{B_{n,m}^2 \log^7(pn)/n\}^{1/6}.
\end{aligned} \tag{25}
$$

Finally, by combining (24) with (25), for any $\mu^X - \mu^Y \in \mathbb{R}^p$, we have that with probability one,

$$
\big|\mathrm{Power}^*(\mu^X - \mu^Y) - \mathrm{Power}(\mu^X - \mu^Y)\big| \lesssim \{B_{n,m}^2 \log^7(pn)/n\}^{1/6},
$$

which completes the proof. □

PROOF OF THEOREM 5. First of all, on the basis of (8) and the triangle inequality, it is clear that

$$
\begin{aligned}
\mathrm{Power}^*(\mu^X - \mu^Y) \ge P_{e^*}\{&\|S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y}\|_\infty \\
&\le \|n^{1/2}(\mu^X - \mu^Y)\|_\infty - c_B(\alpha)\}.
\end{aligned} \tag{26}
$$

At this point, with some abuse of notation, we denote $\{e_j : j \le p\}$ as the natural basis for $\mathbb{R}^p$. Then it follows from union bound inequality and concentration inequality that for any $t \ge 0$,

$$
\begin{aligned}
&P_{e^*}\{\|S_n^{e^*X} - n^{1/2}m^{-1/2}S_m^{e^*Y}\|_\infty \ge t\} \\
&\quad \le \sum_{j=1}^p P_{e^*}\{|S_{nj}^{e^*X} - n^{1/2}m^{-1/2}S_{mj}^{e^*Y}| \ge t\} \\
&\quad \le \sum_{j=1}^p 2\exp[-t^2/\{2e_j'(\hat{\Sigma}^X + nm^{-1}\hat{\Sigma}^Y)e_j\}] \\
&\quad \le 2p\exp\Big(-t^2/\Big[2\max_{j \le p}\{e_j'(\hat{\Sigma}^X + nm^{-1}\hat{\Sigma}^Y)e_j\}\Big]\Big).
\end{aligned} \tag{27}
$$

By plugging $t = c_B(\alpha)$ into (27), it follows from the definition of $c_B(\alpha)$ that

$$
\begin{aligned}
c_B(\alpha) &\le \Big[2\log(2p/\alpha)\max_{j \le p}\{e_j'(\hat{\Sigma}^X + nm^{-1}\hat{\Sigma}^Y)e_j\}\Big]^{1/2} \\
&\le \Big[4\log(pn)\max_{j \le p}\{e_j'(\hat{\Sigma}^X + nm^{-1}\hat{\Sigma}^Y)e_j\}\Big]^{1/2},
\end{aligned} \tag{28}
$$

for sufficiently large n. To bound the quantity $\max_{j\le p}\{e_j'(\hat{\Sigma}^X + nm^{-1}\hat{\Sigma}^Y)e_j\}$, we notice that

$$
\begin{aligned}
&\max_{j\le p}\{e_j'(\hat{\Sigma}^X + nm^{-1}\hat{\Sigma}^Y)e_j\} \\
&\quad = \|\hat{\Sigma}^X + nm^{-1}\hat{\Sigma}^Y\|_\infty \qquad (29)\\
&\quad \le \|\hat{\Sigma}^X - \Sigma^X + nm^{-1}(\hat{\Sigma}^Y - \Sigma^Y)\|_\infty + \|\Sigma^X + nm^{-1}\Sigma^Y\|_\infty.
\end{aligned}
$$

For the term $\|\hat{\Sigma}^X - \Sigma^X + nm^{-1}(\hat{\Sigma}^Y - \Sigma^Y)\|_\infty$, inequalities (53) and (54) from the Supplementary Material together with (12), (17) and condition (a) entails that there exists a universal constant $c_1 > 0$ such that

$$
\|\hat{\Sigma}^X - \Sigma^X + nm^{-1}(\hat{\Sigma}^Y - \Sigma^Y)\|_\infty \le c_1\{B_{n,m}^2 \log^3(pn)/n\}^{1/2}, \qquad (30)
$$

with probability tending to one. Regarding the term $\|\Sigma^X + nm^{-1}\Sigma^Y\|_\infty$, one has

$$
\begin{aligned}
&\|\Sigma^X + nm^{-1}\Sigma^Y\|_\infty \\
&\quad \le \|\Sigma^X\|_\infty + nm^{-1}\|\Sigma^Y\|_\infty \le \|\Sigma^X\|_\infty + c_2\|\Sigma^Y\|_\infty \\
&\quad = \max_{1\le j\le p}\sum_{i=1}^n E\{(X_{ij}-\mu_j^X)^2\}/n + c_2\max_{1\le j\le p}\sum_{i=1}^m E\{(Y_{ij}-\mu_j^Y)^2\}/m \\
&\quad \le \max_{1\le j\le p}\sum_{i=1}^n [E\{(X_{ij}-\mu_j^X)^4\}]^{1/2}/n \qquad (31)\\
&\qquad + c_2\max_{1\le j\le p}\sum_{i=1}^m [E\{(Y_{ij}-\mu_j^Y)^4\}]^{1/2}/m \\
&\quad \le \left[\max_{1\le j\le p}\sum_{i=1}^n E\{(X_{ij}-\mu_j^X)^4\}/n\right]^{1/2} \\
&\qquad + c_2\left[\max_{1\le j\le p}\sum_{i=1}^m E\{(Y_{ij}-\mu_j^Y)^4\}/m\right]^{1/2} \\
&\quad \le c_3 B_{n,m},
\end{aligned}
$$

for some universal constants $c_2, c_3 > 0$, where the second inequality is by condition (a), the third inequality is based on Jensen's inequality, the fourth inequality holds from the Cauchy–Schwarz inequality and the last inequality follows from condition (c). To this end, by combining (30), (31), (e) with (29), it can be deduced that there exists a universal constant $c_4 > 0$ such that

$$
\max_{j\le p}\{e_j'(\hat{\Sigma}^X + nm^{-1}\hat{\Sigma}^Y)e_j\} \le c_4 B_{n,m}, \qquad (32)
$$

with probability tending to one. Together with (28), it can be verified that

$$
c_B(\alpha) \le \{4c_4 B_{n,m}\log(pn)\}^{1/2}, \qquad (33)
$$

with probability tending to one. Now, we set the constant K_s in (f) as $K_s = 4c_4^{1/2}$, and it then follows from (f) and (33) that

$$
\|n^{1/2}(\mu^X - \mu^Y)\|_\infty - c_B(\alpha) \ge \{4c_4 B_{n,m}\log(pn)\}^{1/2}, \qquad (34)
$$

ith obabilit tending to one. Hence, it can be ded ced that ith obabilit tending to one,

$$
\begin{aligned}
&\text{Po e}^{*}(\mu^X - \mu^Y) \\
&\quad \geq P_{e^*}\big[\big\| S_n^{e^* X} - n^{1/2} m^{-1/2} S_m^{e^* Y} \big\|_\infty \leq \{4 c_4 B_{n,m} \log(pn)\}^{1/2}\big] \\
&\quad = 1 - P_{e^*}\big[\big\| S_n^{e^* X} - n^{1/2} m^{-1/2} S_m^{e^* Y} \big\|_\infty \geq \{4 c_4 B_{n,m} \log(pn)\}^{1/2}\big] \\
&\quad \geq 1 - 2p\,\mathrm{ex}\Big(-4 c_4 B_{n,m} \log(pn) \Big/ \Big[2 \max_{j \leq p} \{e_j' (\hat{\Sigma}^X + n m^{-1} \hat{\Sigma}^Y) e_j\Big\}[[[[[([\sum
\end{aligned}
$$

[9] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309–2352. MR3693963 https://doi.org/10.1214/16-AOP1113

[10] FENG, L., ZOU, C., WANG, Z. and ZHU, L. (2015). Two-sample Behrens–Fisher problem for high-dimensional data. *Statist. Sinica* **25** 1297–1312. MR3409068

[11] GREGORY, K. B., CARROLL, R. J., BALADANDAYUTHAPANI, V. and LAHIRI, S. N. (2015). A two-sample test for equality of means in high dimension. *J. Amer. Statist. Assoc.* **110** 837–849. MR3367268 https://doi.org/10.1080/01621459.2014.934826

[12] HU, J., BAI, Z., WANG, C. and WANG, W. (2017). On testing the equality of high dimensional mean vectors with unequal covariance matrices. *Ann. Inst. Statist. Math.* **69** 365–387. MR3611524 https://doi.org/10.1007/s10463-015-0543-8

[13] HUSSAIN, L., AZIZ, W., NADEEM, S. A., SHAH, S. A. and MAJID, A. (2015). Electroencephalography (EEG) analysis of alcoholic and control subjects using multiscale permutation entropy. *J. Multidiscip. Eng. Sci. Technol.* **1** 3159–0040.

[14] PARK, J. and AYYALA, D. N. (2013). A test for the mean vector in large dimension and small samples. *J. Statist. Plann. Inference* **143** 929–943. MR3011304 https://doi.org/10.1016/j.jspi.2012.11.001

[15] SHEN, Y. and LIN, Z. (2015). An adaptive test for the mean vector in large-*p*-small-*n* problems. *Comput. Statist. Data Anal.* **89** 25–38. MR3349665 https://doi.org/10.1016/j.csda.2015.03.004

[16] SRIVASTAVA, M. S. (2007). Multivariate theory for analyzing high dimensional data. *J. Japan Statist. Soc.* **37** 53–86. MR2392485 https://doi.org/10.14490/jjss.37.53

[17] SRIVASTAVA, M. S. (2009). A test for the mean vector with fewer observations than the dimension under non-normality. *J. Multivariate Anal.* **100** 518–532. MR2483435 https://doi.org/10.1016/j.jmva.2008.06.006

[18] SRIVASTAVA, M. S. and DU, M. (2008). A test for the mean vector with fewer observations than the dimension. *J. Multivariate Anal.* **99** 386–402. MR2396970 https://doi.org/10.1016/j.jmva.2006.11.002

[19] SRIVASTAVA, M. S. and KUBOKAWA, T. (2013). Tests for multivariate analysis of variance in high dimension under non-normality. *J. Multivariate Anal.* **115** 204–216. MR3004555 https://doi.org/10.1016/j.jmva.2012.10.011

[20] WANG, L., PENG, B. and LI, R. (2015). A high-dimensional nonparametric multivariate test for mean vector. *J. Amer. Statist. Assoc.* **110** 1658–1669. MR3449062 https://doi.org/10.1080/01621459.2014.988215

[21] XU, G., LIN, L., WEI, P. and PAN, W. (2016). An adaptive two-sample test for high-dimensional means. *Biometrika* **103** 609–624. MR3551787 https://doi.org/10.1093/biomet/asw029

[22] XUE, K. and YAO, F. (2019). Supplement to "Distribution and correlation-free two-sample test of high-dimensional means." https://doi.org/10.1214/19-AOS1848SUPP.

[23] YAGI, A. and SEO, T. (2014). A test for mean vector and simultaneous confidence intervals with three-step monotone missing data. *Amer. J. Math. Management Sci.* **33** 161–175.

[24] YAMADA, T. and HIMENO, T. (2015). Testing homogeneity of mean vectors under heteroscedasticity in high-dimension. *J. Multivariate Anal.* **139** 7–27. MR3349477 https://doi.org/10.1016/j.jmva.2015.02.005

[25] ZHANG, J. and PAN, M. (2016). A high-dimension two-sample test for the mean using cluster subspaces. *Comput. Statist. Data Anal.* **97** 87–97. MR3447038 https://doi.org/10.1016/j.csda.2015.12.004

[26] ZHANG, X. (2015). Testing high dimensional mean under sparsity. Preprint. Available at arXiv:1509.08444v2.

[27] ZHAO, J. (2017). A new test for the mean vector in large dimension and small samples. *Comm. Statist. Simulation Comput.* **46** 6115–6128. MR3740770 https://doi.org/10.1080/03610918.2016.1197244

[28] ZHONG, P.-S., CHEN, S. X. and XU, M. (2013). Tests alternative to higher criticism for high-dimensional means under sparsity and column-wise dependence. *Ann. Statist.* **41** 2820–2851. MR3161449 https://doi.org/10.1214/13-AOS1168

[29] ZHU, Y. and BRADIC, J. (2016). Two-sample testing in non-sparse high-dimensional linear models. Preprint. Available at arXiv:1610.04580v1.