Statistica Sinica **31** (2021), 1853-1869
doi:https://doi.org/10.5705/ss.202019.0170

A ROBUST AND NONPARAMETRIC TWO-SAMPLE TEST IN HIGH DIMENSIONS

Tao Qiu[1], Wangli Xu[1] and Liping Zhu[1,2]

[1]*Renmin University of China and* [2]*Zhejiang Gongshang University*

Abstract: Many tests are available to test the homogeneity of two random samples, that is, the exact equivalence of their statistical distributions. When the two random samples are high dimensional or not normally distributed, the asymptotic null distributions of most existing two-sample tests are rarely tractable. This limits their usefulness in high dimensions, even when the sample sizes are sufficiently large. In addition, existing tests require a careful selection of the tuning parameters to enhance their power performance. However, doing so is very challenging, especially in high dimensions. In this paper, we propose a robust and fully nonparametric two-sample test to detect the heterogeneity of two random samples. Our proposed test is free of tuning parameters. It is built upon the Cramér−von Mises distance, and can be readily used in high dimensions. In addition, our proposed test is robust to the presence of outliers or extreme values in that no moment condition is required. The asymptotic null distribution of our proposed test is standard normal when both the sample sizes and the dimensions of the two random samples diverge to infinity. This facilitates the implementation of our proposed test dramatically, in that no bootstrap or re-sampling technique has to be used to decide an appropriate critical value. We demonstrate the power performance of our proposed test through extensive simulations and real-world applications.

Key words and phrases: Cramér-von Mises test, equality of distributions, high dimension, homogeneity, two-sample test, U-statistics.

1. Introduction

Testing the homogeneity of two independent random samples is one of the most fundamental problems in statistics (Lehmann and Romano (2005); Thas (2010)). Suppose $\{\mathbf{x}_i, i = 1, \ldots, m\}$ and $\{\mathbf{y}_i, i = 1, \ldots, n\}$ are two random samples drawn independently from F and G, respectively. Testing their homogeneity amounts to testing the exact equivalence of their respective distribution functions. In symbols, the two-sample test checks the following:

$$H_0 : F = G \quad \text{versus} \quad H_1 : F \neq G. \tag{1.1}$$

Corresponding author: Liping Zhu, Center for Applied Statistics, Institute of Statistics and Big Data, Renmin University of China, Beijing 100872, China. E-mail: zhu.liping@ruc.edu.cn.

Rejecting H_0 indicates the presence of heterogeneity.

Many tests have been proposed in the literature to test (1.1). In the univariate case of $p = 1$, the Kolmogorov−Smirnov (Smirnov (1939)) and Cramér−von Mises (Rosenblatt (1952); Anderson (1962)) tests are perhaps two of the most popular omnibus tests. Both quantify the discrepancies between the empirical distributions of the two random samples. In the multivariate case of $p \geq 2$, Friedman and Rafsky (1979) and Biswas et al. (2014) proposed homogeneity tests using the minimal spanning tree and the shortest Hamiltonian path, respectively. The tests proposed by Henze (1988), Schilling (1986), and Mondal, Biswas and Ghosh (2015) are all based on the nearest neighbors. Hall and Tajvidi (2002) introduced a permutation test based on ranking the pooled samples. Rosenbaum (2005) devised a run test that matches observations into disjoint pairs. Gretton et al. (2012) introduced a class of distances between two probability distributions in a reproducing kernel Hilbert space, called the maximum mean discrepancy (MMD). However, the above tests require a careful selection of the tuning parameters, such as the weight functions, number of neighbors, or bandwidths of the Gaussian MMD. The power performance of these tests relies heavily on the tuning parameters; however, the optimal selection of these parameters is not straightforward, especially in high dimensions. In addition, Chen and Friedman (2017) pointed out that none of these tests are sensitive to both location shifts and scale differences. Baringhaus and Franz (2004) and Biswas and Ghosh (2014) used the energy distances between the empirical characteristic functions of the two random samples. These two tests require that the second moments of both random samples be finite, and thus are not powerful in the presence of outliers or extreme values. Pan et al. (2018) introduced ball divergence to measure the difference between the empirical probability density functions of the two random samples. Ball divergence requires no moment condition and is free of tuning parameters. However, the asymptotic null distribution of this test is not tractable, even when the dimensions of both random samples are small. Our limited simulations indicate that the ball divergence test is highly insensitive to location shifts. The asymptotic properties of the above tests are also unknown in extremely high dimensions.

In this paper, we propose a robust and fully nonparametric two-sample test. Suppose $\{\mathbf{x}_i = (X_{i1}, \ldots, X_{ip})^{\mathrm{T}}, i = 1, \ldots, m\}$, $\{\mathbf{y}_i = (Y_{i1}, \ldots, Y_{ip})^{\mathrm{T}}, i = 1, \ldots, n\}$, and $\{\mathbf{z}_r = (Z_{r1}, \ldots, Z_{rp})^{\mathrm{T}}, r = 1, \ldots, m + n\}$ are three random samples drawn independently from F, G, and H, respectively, where $H \overset{\text{def}}{=} \{m/(m + n)\}F + \{n/(m + n)\}G$. Denote the distribution functions of X_{ik}, Y_{jk}, and Z_{rk}, by F_k, G_k, and $H_k \overset{\text{def}}{=} \{m/(m + n)\}F_k + \{n/(m + n)\}G_k$, respectively, for $k = 1, \ldots, p$.

In the present context, we consider testing

$$H_0 : F_k = G_k \text{ for all } 1 \le k \le p, \text{ versus}$$
$$H_1 : F_k \neq G_k \text{ for at least one } k \in \{1, \ldots, p\}. \tag{1.2}$$

The exact equivalence between F and G is not fully characterized by the distances between F_k and G_k, for $k = 1, \ldots, p$. However, we argue that, in many real-world problems the distances between F_k and G_k are usually informative in testing the exact equivalence between F and G.

We propose quantifying the degree of deviation from H_0 in (1.2) using the U-statistic estimate of

$$Q \stackrel{\text{def}}{=} \sum_{k=1}^{p} \left[\int_{-\infty}^{\infty} \{F_k(z) - G_k(z)\}^2 dH_k(z) \right].$$

The quantity Q is based on the Cramér−von Mises distance, and can be readily used in arbitrarily high dimensions. We propose estimating the distribution functions F_k, G_k, and H_k, using their corresponding empirical distribution functions $\widehat{F}_k$, $\widehat{G}_k$, and $\widehat{H}_k \stackrel{\text{def}}{=} \{m/(m+n)\}\widehat{F}_k + \{n/(m+n)\}\widehat{G}_k$, respectively. Consequently, the U-statistic estimate of Q is free of tuning parameters and is robust to the presence of outliers and extreme values in either of the two random samples. We advocate using Q for at least two additional reasons. First, this allows for arbitrarily large p. The computational complexity for estimating Q is linear in p. Second, the asymptotic null distribution is standard normal, regardless of the relationship between p and $\min(m, n)$. Therefore, no re-sampling or bootstrap procedure has to be used to approximate the asymptotic null distribution. These two properties facilitate the implementation of our proposed test in extremely high dimensions, and allow us to handle very large-scale data sets. The computation complexity of comparing F and G directly in high-dimensional problems is prohibitive.

In Section 2, we give an explicit form for the U-statistic estimate of Q. This allows us to use the Hoeffding decomposition and martingale central limit theorem to derive the asymptotic properties of our proposed test. Extensive numerical studies are conducted in Section 3 to demonstrate the power performance of our proposed test, and to compare it with that of several existing tests. Our empirical studies indicate that the proposed two-sample test is sensitive to both location shifts and scale differences, even in high dimensions. We conclude this paper with a brief discussion in Section 4. All technical details are relegated to the Supplementary Material.

2. The Test Procedure

In this section, we introduce our proposed two-sample test.

2.1. The U-statistic estimate of Q

We assume throughout that Z_{rk} is independent of X_{ik} and Y_{jk} and drawn independently from H_k, for $k = 1, \ldots, p$, $i = 1, \ldots, m$, $j = 1, \ldots, n$, and $r = 1, \ldots, m+n$. Then, an equivalent form of Q is given by

$$Q = \sum_{k=1}^{p} E\Big[\big\{I(X_{1k} \le Z_{1k}) - I(Y_{1k} \le Z_{1k})\big\}\big\{I(X_{2k} \le Z_{1k}) - I(Y_{2k} \le Z_{1k})\big\}\Big],$$

where the expectation E is taken with respect to F_k, G_k, and H_k. Define

$$\begin{aligned}
Q_1 &\stackrel{\text{def}}{=} \sum_{k=1}^{p} E\left\{I(X_{1k} \le Z_{1k}) I(X_{2k} \le Z_{1k})\right\}, \\
Q_2 &\stackrel{\text{def}}{=} \sum_{k=1}^{p} E\left\{I(Y_{1k} \le Z_{1k}) I(Y_{2k} \le Z_{1k})\right\}, \text{and} \\
Q_3 &\stackrel{\text{def}}{=} -2\sum_{k=1}^{p} E\left\{I(X_{1k} \le Z_{1k}) I(Y_{2k} \le Z_{1k})\right\}.
\end{aligned}$$

It can be verified that $Q = Q_1 + Q_2 + Q_3$. In the above definitions, the summands are of similar forms. To simplify subsequent illustrations, define

$$\rho(\mathbf{x}_i, \mathbf{x}_j, \mathbf{z}_r) \stackrel{\text{def}}{=} \sum_{k=1}^{p} I(X_{ik} \le Z_{rk}) I(X_{jk} \le Z_{rk}).$$

The U-statistic estimates of Q_1, Q_2, and Q_3 are defined, respectively, by

$$\begin{aligned}
\widehat{Q}_1 &\stackrel{\text{def}}{=} \{m(m-1)(m+n)\}^{-1} \sum_{i \ne j}^{m} \sum_{r=1}^{m+n} \rho(\mathbf{x}_i, \mathbf{x}_j, \mathbf{z}_r), \\
\widehat{Q}_2 &\stackrel{\text{def}}{=} \{n(n-1)(m+n)\}^{-1} \sum_{i \ne j}^{n} \sum_{r=1}^{m+n} \rho(\mathbf{y}_i, \mathbf{y}_j, \mathbf{z}_r), \quad \text{and} \\
\widehat{Q}_3 &\stackrel{\text{def}}{=} -2\left\{mn(m+n)\right\}^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{r=1}^{m+n} \rho(\mathbf{x}_i, \mathbf{y}_j, \mathbf{z}_r).
\end{aligned}$$

Both $\widehat{Q}_1$ and $\widehat{Q}_2$ are two-sample U-statistics of order $(2;1)$, and $\widehat{Q}_3$ is a three-sample U-statistic of order $(1;1;1)$. Define

$$\widehat{Q} \stackrel{\text{def}}{=} \widehat{Q}_1 + \widehat{Q}_2 + \widehat{Q}_3, \tag{2.1}$$

that is, the U-statistic estimate of Q.

2.2. Notation

The following notation is used in subsequent expositions. Define $U_k(X_{ik}, Z_{rk}) \stackrel{\text{def}}{=} I(X_{ik} \le Z_{rk}) - F_k(Z_{rk})$ and $V_k(Y_{ik}, Z_{rk}) \stackrel{\text{def}}{=} I(Y_{ik} \le Z_{rk}) - G_k(Z_{rk})$. We further define

$$\omega_{11}(\mathbf{x}_i, \mathbf{x}_j, \mathbf{z}_r) \stackrel{\text{def}}{=} \sum_{k=1}^{p} U_k(X_{ik}, Z_{rk}) U_k(X_{jk}, Z_{rk}),$$

$$\omega_{12}(\mathbf{y}_i, \mathbf{y}_j, \mathbf{z}_r) \stackrel{\text{def}}{=} \sum_{k=1}^{p} V_k(Y_{ik}, Z_{rk}) V_k(Y_{jk}, Z_{rk}),$$

$$\text{and} \quad \omega_{13}(\mathbf{x}_i, \mathbf{y}_j, \mathbf{z}_r) \stackrel{\text{def}}{=} \sum_{k=1}^{p} U_k(X_{ik}, Z_{rk}) V_k(Y_{jk}, Z_{rk}).$$

Based on the above notation, we define

$$\begin{aligned} \widehat{T}_1 \stackrel{\text{def}}{=} & \{m(m-1)(m+n)\}^{-1} \sum_{i \neq j}^{m} \sum_{r=1}^{m+n} \omega_{11}(\mathbf{x}_i, \mathbf{x}_j, \mathbf{z}_r) + \{n(m+n)\}^{-1} \\ & \left\{ (n-1)^{-1} \sum_{i \neq j}^{n} \sum_{r=1}^{m+n} \omega_{12}(\mathbf{y}_i, \mathbf{y}_j, \mathbf{z}_r) - 2m^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{r=1}^{m+n} \omega_{13}(\mathbf{x}_i, \mathbf{y}_j, \mathbf{z}_r) \right\}. \end{aligned} \tag{2.2}$$

Here, $\widehat{T}_1$ has a complicated form, in that it is a U-statistic estimate of three random samples. We further define

$$\begin{aligned} \widehat{T}_{1,1} \stackrel{\text{def}}{=} & \{m(m-1)\}^{-1} \sum_{i \neq j}^{m} \varphi_{11}(\mathbf{x}_i, \mathbf{x}_j) + \{n(n-1)\}^{-1} \sum_{i \neq j}^{n} \varphi_{12}(\mathbf{y}_i, \mathbf{y}_j) \\ & -2(mn)^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n} \varphi_{13}(\mathbf{x}_i, \mathbf{y}_j), \end{aligned} \tag{2.3}$$

which is a two-sample U-statistic, where

$$\varphi_{11}(\mathbf{x}_i, \mathbf{x}_j) \stackrel{\text{def}}{=} \sum_{k=1}^{p} E\left\{ U_k(X_{ik}, Z_{rk}) U_k(X_{jk}, Z_{rk}) \mid X_{ik}, X_{jk}, Y_{ik}, Y_{jk} \right\},$$

$$\varphi_{12}(\mathbf{y}_i, \mathbf{y}_j) \stackrel{\text{def}}{=} \sum_{k=1}^{p} E\left\{ V_k(Y_{ik}, Z_{rk}) V_k(Y_{jk}, Z_{rk}) \mid X_{ik}, X_{jk}, Y_{ik}, Y_{jk} \right\}, \text{ and}$$

$$\varphi_{13}(\mathbf{x}_i, \mathbf{y}_j) \overset{\text{def}}{=} \sum_{k=1}^{p} E\left\{U_k(X_{ik}, Z_{rk})V_k(Y_{jk}, Z_{rk}) \mid X_{ik}, X_{jk}, Y_{ik}, Y_{jk}\right\}. \tag{2.4}$$

Let $D_k(Z_{rk}) \overset{\text{def}}{=} F_k(Z_{rk}) - G_k(Z_{rk})$. Define

$$\begin{aligned} \omega_{21}(\mathbf{x}_i, \mathbf{z}_r) &\overset{\text{def}}{=} \sum_{k=1}^{p} D_k(Z_{rk})U_k(X_{ik}, Z_{rk}), \\ \omega_{22}(\mathbf{y}_j, \mathbf{z}_r) &\overset{\text{def}}{=} \sum_{k=1}^{p} D_k(Z_{rk})V_k(Y_{jk}, Z_{rk}), \\ \text{and} \qquad \omega_{23}(\mathbf{z}_r) &\overset{\text{def}}{=} \sum_{k=1}^{p} D_k^2(Z_{rk}). \end{aligned} \tag{2.5}$$

Based on the above notation, we define

$$\begin{aligned} \widehat{T}_2 \overset{\text{def}}{=} (m+n)^{-1} \Bigg[& 2m^{-1} \sum_{i=1}^{m} \sum_{r=1}^{m+n} \omega_{21}(\mathbf{x}_i, \mathbf{z}_r) - 2n^{-1} \sum_{j=1}^{n} \sum_{r=1}^{m+n} \omega_{22}(\mathbf{y}_j, \mathbf{z}_r) \\ & + \sum_{r=1}^{m+n} \{\omega_{23}(\mathbf{z}_r) - D_0\} \Bigg], \end{aligned} \tag{2.6}$$

where

$$D_0 \overset{\text{def}}{=} \sum_{k=1}^{p} E\{D_k^2(Z_{rk})\} = E\left\{\omega_{23}(\mathbf{z}_r)\right\}. \tag{2.7}$$

Note that both $\widehat{T}_2$ and D_0 quantify deviations from H_0 in (1.2). Specifically, under H_0 in (1.2), $D_k(Z_{rk}) = 0$; accordingly, $D_0 = 0$, which, together with (2.5) and (2.6), yields that $\widehat{T}_2 = 0$.

2.3. The asymptotic behavior of $\widehat{Q}$

We outline the asymptotic behavior of $\widehat{Q}$ in Section 2.3, and establish these properties rigorously in Section 2.4.

We use the Hoeffding decomposition and martingale central limit theorem to analyze the asymptotic behavior of $\widehat{Q}$. However, the expectations of $\widehat{Q}_1$, $\widehat{Q}_2$, and $\widehat{Q}_3$ are all nonzero under H_0 in (1.2). To facilitate subsequent asymptotic derivations, we rewrite $\widehat{Q}$ as $\widehat{Q} = \widehat{T}_1 + \widehat{T}_2 + D_0$, where $\widehat{T}_1$ and $\widehat{T}_2$ are given in (2.2) and (2.6), respectively, and D_0 is a constant defined in (2.7). By the Hoeffding decomposition, $\widehat{T}_1$ can be approximated precisely using the two-sample U-statistic $\widehat{T}_{1,1}$ in (2.3). In symbols, $\widehat{T}_1 = \widehat{T}_{1,1}\{1 + o_p(1)\}$. In addition, $\widehat{T}_2 = D_0 = 0$ holds exactly under H_0 in (1.2), indicating that $\widehat{T}_2$ and D_0 quantify the degree of heterogeneity of the two random samples.

The three summands on the right-hand side of $\widehat{T}_{1,1}$ in (2.3) have zero mean identically and are uncorrelated with each other. Therefore, the asymptotic variance of $\widehat{T}_{1,1}$ can be derived without much difficulty. To be precise, let $\sigma_{11}^2 \stackrel{\text{def}}{=} E\{\varphi_{11}^2(\mathbf{x}_1, \mathbf{x}_2)\}$, $\sigma_{12}^2 \stackrel{\text{def}}{=} E\{\varphi_{12}^2(\mathbf{y}_1, \mathbf{y}_2)\}$, and $\sigma_{13}^2 \stackrel{\text{def}}{=} E\{\varphi_{13}^2(\mathbf{x}, \mathbf{y})\}$. It follows that

$$\operatorname{var}(\widehat{T}_{1,1}) = 2\{m(m-1)\}^{-1}\sigma_{11}^2 + 2\{n(n-1)\}^{-1}\sigma_{12}^2 + 4(mn)^{-1}\sigma_{13}^2. \quad (2.8)$$

In Section 2.4, we establish the asymptotic normality of $\widehat{T}_{1,1}$ using the martingale central limit theorem. In addition, under H_0 in (1.2) and mild regularity conditions, $\widehat{Q}/\{\operatorname{var}(\widehat{T}_{1,1})\}^{1/2}$ is asymptotically standard normal as m, n, and p diverge to ∞. Therefore, as long as $\operatorname{var}(\widehat{T}_{1,1})$ is estimated consistently, the distribution of $\widehat{Q}$ is asymptotically tractable.

In what follows, we provide an estimate of $\operatorname{var}(\widehat{T}_{1,1})$. We define the leave-one-observation-out estimates of F_k and G_k, respectively, by

$$\widehat{F}_{k(-i)}(Z_{rk}) \stackrel{\text{def}}{=} (m-1)^{-1}\sum_{l\neq i}^{m} I(X_{lk} \leq Z_{rk}),$$
$$\text{and} \quad \widehat{G}_{k(-j)}(Z_{rk}) \stackrel{\text{def}}{=} (n-1)^{-1}\sum_{l\neq j}^{n} I(Y_{lk} \leq Z_{rk}).$$

Similarly, we define the leave-two-observations-out estimates as

$$\widehat{F}_{k(-i,-j)}(Z_{rk}) \stackrel{\text{def}}{=} (m-2)^{-1}\sum_{l\neq i, l\neq j}^{m} I(X_{lk} \leq Z_{rk}) \text{ and}$$
$$\widehat{G}_{k(-i,-j)}(Z_{rk}) \stackrel{\text{def}}{=} (n-2)^{-1}\sum_{l\neq i, l\neq j}^{n} I(Y_{lk} \leq Z_{rk}).$$

Instead of using the classic empirical distributions $\widehat{F}_k$ and $\widehat{G}_k$ directly, we use $\widehat{F}_{k(-i)}$, $\widehat{G}_{k(-i)}$, $\widehat{F}_{k(-i,-j)}$, and $\widehat{G}_{k(-i,-j)}$ to estimate the asymptotic variance. This yields an unbiased estimate of $\operatorname{var}(\widehat{T}_{1,1})$. The empirical studies in Section 3 indicate that the bias is substantial if we use $\widehat{F}_k$ and $\widehat{G}_k$ to estimate $\operatorname{var}(\widehat{T}_{1,1})$. This echoes the observations of many other studies on high dimensions. See, for example, Chen and Qin (2010), Zhong and Chen (2011), and Zhang, Yao and Shao (2018).

We further define

$$\widehat{\sigma}_{11}^2 \stackrel{\text{def}}{=} \left\{4\binom{m}{2}\binom{m+n}{2}\right\}^{-1} \sum_{i\neq j}^{m}\sum_{r\neq s}^{m+n} \left\{\left[\sum_{k=1}^{p}\{I(X_{ik} \le Z_{rk}) - \widehat{F}_{k(-i,-j)}(Z_{rk})\}\right.\right.$$
$$\left.\left. I(X_{jk} \le Z_{rk})\right]\left[\sum_{k=1}^{p} I(X_{ik} \le Z_{sk})\{I(X_{jk} \le Z_{sk}) - \widehat{F}_{k(-i,-j)}(Z_{sk})\}\right]\right\},$$
$$\widehat{\sigma}_{12}^2 \stackrel{\text{def}}{=} \left\{4\binom{n}{2}\binom{m+n}{2}\right\}^{-1} \sum_{i\neq j}^{n}\sum_{r\neq s}^{m+n} \left\{\left[\sum_{k=1}^{p}\{I(Y_{ik} \le Z_{rk}) - \widehat{G}_{k(-i,-j)}(Z_{rk})\}\right.\right.$$
$$\left.\left. I(Y_{jk} \le Z_{rk})\right]\left[\sum_{k=1}^{p} I(Y_{ik} \le Z_{sk})\{I(Y_{jk} \le Z_{sk}) - \widehat{G}_{k(-i,-j)}(Z_{sk})\}\right]\right\}, \text{ and,}$$
$$\widehat{\sigma}_{13}^2 \stackrel{\text{def}}{=} \left\{2mn\binom{m+n}{2}\right\}^{-1} \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{r\neq s}^{m+n} \left\{\left[\sum_{k=1}^{p}\{I(X_{ik} \le Z_{rk}) - \widehat{F}_{k(-i)}(Z_{rk})\}\right.\right.$$
$$\left.\left. I(Y_{jk} \le Z_{rk})\right]\left[\sum_{k=1}^{p} \{I(Y_{jk} \le Z_{sk}) - \widehat{G}_{k(-j)}(Z_{sk})\}I(X_{ik} \le Z_{sk})\right]\right\}.$$

The unbiased estimate of $\text{var}(\widehat{T}_{1,1})$ is thus given by

$$\widehat{\text{var}}(\widehat{T}_{1,1}) \stackrel{\text{def}}{=} 2\{m(m-1)\}^{-1}\widehat{\sigma}_{11}^2 + 2\{n(n-1)\}^{-1}\widehat{\sigma}_{12}^2 + 4(mn)^{-1}\widehat{\sigma}_{13}^2.$$

To implement the test, we can reject H_0 soundly at the significance level α as long as the test statistic, $\widehat{Q}^2/\widehat{\text{var}}(\widehat{T}_{1,1})$, is greater than or equal to $z_{1-\alpha/2}^2$, where $z_{1-\alpha/2}$ denotes the $(1-\alpha/2) \times 100\%$th quantile of the standard normal distribution.

2.4. The asymptotic behavior of $\widehat{Q}/\{\widehat{\text{var}}(\widehat{T}_{1,1})\}^{1/2}$

We rigorously study the asymptotic behavior of the test statistic $\widehat{Q}/\left\{\widehat{\text{var}}(\widehat{T}_{1,1})\right\}^{1/2}$ under H_0 and H_1. Throughout, we assume

(C1) $m/(m+n) \to c \in (0,1)$ as both m and n diverge to infinity.

Condition (C1) is commonly assumed in two-sample tests. Along with (C1), we assume the following conditions. Define $\nu_1 \stackrel{\text{def}}{=} \sigma_{11}^{-2}E\{\omega_{11}^2(\mathbf{x}_1,\mathbf{x}_2,\mathbf{z})\} + \sigma_{12}^{-2} E\{\omega_{12}^2(\mathbf{y}_1,\mathbf{y}_2,\mathbf{z})\} + \sigma_{13}^{-2}E\{\omega_{13}^2(\mathbf{x},\mathbf{y},\mathbf{z})\}$. Assume that

(C2) $\nu_1/m \to 0$, as $p \to \infty$.

Condition (C2) ensures that $\widehat{T}_1 = \widehat{T}_{1,1}\{1+o_p(1)\}$, which is satisfied when

$$E\{\omega_{11}^2(\mathbf{x}_1,\mathbf{x}_2,\mathbf{z})\} = o(m\sigma_{11}^2), \tag{2.9}$$

because $E\{\omega_{11}^2(\mathbf{x}_1,\mathbf{x}_2,\mathbf{z})\} = E\{\omega_{12}^2(\mathbf{y}_1,\mathbf{y}_2,\mathbf{z})\} = E\{\omega_{13}^2(\mathbf{x},\mathbf{y},\mathbf{z})\}$ and $\sigma_{11}^2 = \sigma_{12}^2 = \sigma_{13}^2$ under H_0 in (1.2). Next, we explore the conditions under which (2.9) holds. Define $h(Z_{1k}, Z_{2l}) \stackrel{\text{def}}{=} \mathrm{cov}\{I(X_{1k} \le Z_{1k}), I(X_{1l} \le Z_{2l}) \mid Z_{1k}, Z_{2l}\}$ and

$$H_{k,l} \stackrel{\text{def}}{=} \left[\mathrm{var}\left\{h(Z_{1k}, Z_{2l})\right\} + E^2\left\{h(Z_{1k}, Z_{2l})\right\}\right]^{1/2}. \tag{2.10}$$

We further write $\mathbf{H} \stackrel{\text{def}}{=} (H_{k,l})_{p\times p}$. Simple algebraic calculations show that

$$\sigma_{11}^2 = E\{\varphi_{11}^2(\mathbf{x}_1,\mathbf{x}_2)\} = \mathrm{tr}(\mathbf{H}^2), \text{ and } E\{\omega_{11}^2(\mathbf{x}_1,\mathbf{x}_2,\mathbf{z})\} = O\left\{\mathrm{tr}(\mathbf{H}^2)\right\}, \tag{2.11}$$

where $\mathrm{tr}(\mathbf{H}^2)$ denotes the trace of $\mathbf{H}^2$, which includes (2.9) directly. Therefore, Condition (C2) is satisfied naturally under H_0. Following similar arguments, we can show that (C2) can also be easily met under H_1. In other words, Condition (C2) is very mild, though this is not very intuitive.

Define $\nu_2 \stackrel{\text{def}}{=} \sigma_{11}^{-4}E[E^2\{\varphi_{11}^2(\mathbf{x}_1,\mathbf{x}_2) \mid \mathbf{x}_1\}] + \sigma_{12}^{-4}E[E^2\{\varphi_{12}^2(\mathbf{y}_1,\mathbf{y}_2) \mid \mathbf{y}_1\}] + \sigma_{13}^{-4}E[E^2\{\varphi_{13}^2(\mathbf{x},\mathbf{y}) \mid \mathbf{x}\}]$, $\nu_3 \stackrel{\text{def}}{=} \sigma_{11}^{-4}E[E^2\{\varphi_{11}(\mathbf{x}_1,\mathbf{x}_2)\varphi_{11}(\mathbf{x}_1,\mathbf{x}_3) \mid \mathbf{x}_2,\mathbf{x}_3\}] + \sigma_{12}^{-4}E[E^2\{\varphi_{12}(\mathbf{y}_1,\mathbf{y}_2)\varphi_{12}(\mathbf{y}_1,\mathbf{y}_3) \mid \mathbf{y}_2,\mathbf{y}_3\}] + \sigma_{13}^{-4}E[E^2\{\varphi_{13}(\mathbf{x}_1,\mathbf{y}_1)\varphi_{13}(\mathbf{x}_1,\mathbf{y}_2) \mid \mathbf{y}_1,\mathbf{y}_2\}]$, and $\nu_4 \stackrel{\text{def}}{=} \sigma_{11}^{-4}E\{\varphi_{11}^4(\mathbf{x}_1,\mathbf{x}_2)\} + \sigma_{12}^{-4}E\{\varphi_{12}^4(\mathbf{y}_1,\mathbf{y}_2)\} + \sigma_{13}^{-4}E\{\varphi_{13}^4(\mathbf{x},\mathbf{y})\}$. We further make the following two assumptions:

(C3) $\nu_3 \to 0$, as $p \to \infty$.

(C4) $(\nu_2 + \nu_4/m)/m \to 0$, as $p \to \infty$.

Conditions (C3)$-$(C4) ensure that $\widehat{T}_{1,1}$ is asymptotically normal. Note that, under H_0 in (1.2), Condition (C3) reduces to

$$E[E^2\{\varphi_{11}(\mathbf{x}_1,\mathbf{x}_2)\varphi_{11}(\mathbf{x}_1,\mathbf{x}_3) \mid \mathbf{x}_2,\mathbf{x}_3\}] = o(\sigma_{11}^4),$$

and Condition (C4) reduces to

$$E[E^2\{\varphi_{11}^2(\mathbf{x}_1,\mathbf{x}_2) \mid \mathbf{x}_1\}] = o(m\sigma_{11}^4), \text{ and } E\{\varphi_{11}^4(\mathbf{x}_1,\mathbf{x}_2)\} = o(m^2\sigma_{11}^4).$$

Similar equalities apply to the $\mathbf{y}$ sample. The above conditions hold if the correlation matrices of $\mathbf{x} = (X_1,\ldots,X_p)^{\mathrm{T}}$ and $\mathbf{y} = (Y_1,\ldots,Y_p)^{\mathrm{T}}$ have a correlated or a banded dependence structure. The definitions of correlated and banded dependence structures are provided in Appendix B.

Direct algebraic calculations show that

$$E\{\varphi_{11}^4(\mathbf{x}_1,\mathbf{x}_2)\} = O\{\mathrm{tr}^2(\mathbf{H}^2)\}, \; E[E^2\{\varphi_{11}^2(\mathbf{x}_1,\mathbf{x}_2) \mid \mathbf{x}_1\}] = O\{\mathrm{tr}^2(\mathbf{H}^2)\}, \tag{2.12}$$

and

$$E[E^2\{\varphi_{11}(\mathbf{x}_1,\mathbf{x}_2)\varphi_{11}(\mathbf{x}_1,\mathbf{x}_3) \mid \mathbf{x}_2,\mathbf{x}_3\}] = O\left\{\mathrm{tr}(\mathbf{H}^4)\right\}. \tag{2.13}$$

This immediately yields Condition (C4) under H_0. In addition, Condition (C3) is implied by $\mathrm{tr}(\mathbf{H}^4) = o\left\{\mathrm{tr}^2(\mathbf{H}^2)\right\}$. Similar assumptions are used in the literature. See, for example, condition (3.8) in Chen and Qin (2010), and as a sufficient condition for Theorem 2.1 in Zhang, Yao and Shao (2018). Detailed derivations of (2.11)−(2.13) are relegated to the Supplementary Material.

Theorem 1. *Under* (C1)–(C4), $\widehat{Q}/\left\{var(\widehat{T}_{1,1})\right\}^{1/2}$ *is asymptotically standard normal under* H_0 *in* (1.2) *as both* p *and* $\min(m,n)$ *diverge to* ∞.

The following theorem states the consistency of $\widehat{\mathrm{var}}(\widehat{T}_{1,1})$.

Theorem 2. *Under* (C1)–(C4), $\widehat{var}(\widehat{T}_{1,1})$ *converges in probability to* $var(\widehat{T}_{1,1})$ *as both* p *and* $\min(m,n)$ *diverge to* ∞.

It follows immediately from Slutsky's theorem that $\widehat{Q}/\left\{\widehat{\mathrm{var}}(\widehat{T}_{1,1})\right\}^{1/2}$ is asymptotically standard normal under H_0.

Recall the definitions of ω_{21}, ω_{22}, and ω_{23} given in (2.5). We define

$$\begin{aligned}\nu_5 \stackrel{\text{def}}{=} n^{-2}\Big(&E\{\omega_{21}^2(\mathbf{x},\mathbf{z})\} + E\{\omega_{22}^2(\mathbf{y},\mathbf{z})\} + nE[E^2\{\omega_{21}(\mathbf{x},\mathbf{z}) \mid \mathbf{x}\}] \\ &+ nE[E^2\{\omega_{21}(\mathbf{y},\mathbf{z}) \mid \mathbf{y}\}] + n\mathrm{var}\{\omega_{23}(\mathbf{z})\}\Big).\end{aligned}$$

We study the power performance under the local alternative $H_1' : \nu_5 = o\left\{\mathrm{var}(\widehat{T}_{1,1})\right\}$. It can be verified that, under H_1',

$$\max_{1\le k\le p} E\{F_k(Z_{rk}) - G_k(Z_{rk})\}^2 = o_p(n^{-1/2}),$$

indicating that H_1' does not deviate from H_0 substantially. Under H_1', the asymptotic variance of $\widehat{Q}$ remains unchanged, which is formally stated in Theorem 3. In general, the asymptotic variance of $\widehat{Q}$ is inflated under the fixed alternative H_1, which would lead to unstable performance of our proposed test. Therefore, we investigate the power performance of our test under the local alternative H_1' only.

Theorem 3. *Under* (C1)–(C4), $(\widehat{Q} - D_0)/\left\{\widehat{var}(\widehat{T}_{1,1})\right\}^{1/2}$ *is asymptotically standard normal under* H_1' *as both* p *and* $\min(m,n)$ *diverge to* ∞, *where* D_0 *is defined in* (2.7).

The power under the local alternative H_1' is given by

$$1-\beta \stackrel{\text{def}}{=} \text{pr}\left\{\frac{\widehat{Q}^2}{\widehat{\text{var}}(\widehat{T}_{1,1})} \geq z^2_{1-\alpha/2}\right\} \to 1-\text{pr}\left\{\chi^2(1) \leq z^2_{1-\alpha/2} - D_0\frac{2\widehat{Q}-D_0}{\widehat{\text{var}}(\widehat{T}_{1,1})}\right\}, \tag{2.14}$$

where $\chi^2(1)$ denotes a χ^2 random variable with one degree of freedom.

An important implication of (2.14) is that the power of our proposed test is largely determined by $D_0(2\widehat{Q} - D_0)/\widehat{\text{var}}(\widehat{T}_{1,1})$. Recall that $\widehat{Q} = \widehat{T}_1 + \widehat{T}_2 + D_0$. Theorem 1 indicates that $\widehat{T}_1$ is asymptotically normal with mean zero, and Theorem 2 ensures that $\widehat{T}_2$ is asymptotically negligible. Therefore, $\widehat{Q}$ converges in probability to D_0 and, accordingly, $D_0(2\widehat{Q} - D_0)$ converges in probability to D_0^2. In addition, by the definition of $\widehat{\text{var}}(\widehat{T}_{1,1})$ given in (2.9), $\widehat{\text{var}}(\widehat{T}_{1,1}) = O_p(p^2/n^2)$, along with Condition (C1). Consequently, $D_0(2\widehat{Q} - D_0)/\widehat{\text{var}}(\widehat{T}_{1,1})$ is asymptotically of order $O_p(n^2D_0^2/p^2)$. Recall that

$$D_0 = \sum_{k=1}^{p} E\{F_k(Z_{rk}) - G_k(Z_{rk})\}^2.$$

If $F_k \neq G_k$ for most k, then it is reasonable to expect D_0 to be a large number of order p. In this case, the power of our proposed test approaches one, asymptotically. However, if $F_k = G_k$ for most k, it is natural to expect D_0 to be a very small number. In this case, our proposed test may suffer from low power performance, unless the sample size $\min(m, n)$ is sufficiently large and the dimension p is relatively small.

3. Numerical Studies

We conduct numerical studies to demonstrate the finite-sample performance of our proposed test, and to compare it with that of two-sample tests in the literature. Existing tests can be classified into three classes. In the first class, tuning parameters need to be specified carefully. Typical examples include Henze (1988), Mondal, Biswas and Ghosh (2015), Biswas et al. (2014), and Hall and Tajvidi (2002). To ease subsequent illustration, we refer to these tests as H, MBG, BMG, and HT, respectively. In particular, the first three tests require specifying the number of nearest neighbors. Following Hall and Tajvidi (2002), we choose $\gamma = 2$ and $w_1(j) = w_2(j) = 1$ in the HT test. In the second class, moment conditions are required to ensure the existence of energy distances. Examples include Baringhaus and Franz (2004), Rosenbaum (2005), and Biswas and Ghosh (2014), which we refer to as BF, R, and BG, respectively. We also include Pan et al. (2018) in our numerical comparison, which belongs to the third class, in which no moment condition or tuning parameter is required. We refer to this test

Table 1. The empirical sizes and powers when $d = 2$.

	QXZ	PTWZ	R	BMG	BG	BF	HT	H	MBG
p	$(\delta, \sigma^2) = (0, 1.0)$								
30	0.045	0.052	0.051	0.050	0.052	0.044	0.052	0.051	0.042
500	0.047	0.044	0.058	0.042	0.064	0.067	0.062	0.056	0.052
2,000	0.037	0.048	0.061	0.053	0.052	0.054	0.044	0.035	0.055
p	$(\delta, \sigma^2) = (0.25, 1.0)$								
30	0.193	0.053	0.147	0.102	0.052	0.141	0.005	0.236	0.077
500	0.658	0.045	0.759	0.500	0.051	0.189	0.058	0.773	0.067
2,000	0.801	0.049	0.995	0.885	0.049	0.191	0.058	0.825	0.043
p	$(\delta, \sigma^2) = (0.15, 2.0)$								
30	0.354	0.537	0.099	0.129	0.289	0.424	0.365	0.315	0.604
500	0.760	0.586	0.216	0.358	0.302	0.539	0.374	0.316	0.561
2,000	0.785	0.614	0.427	0.621	0.320	0.551	0.374	0.341	0.489
p	$(\delta, \sigma^2) = (0.0, 2.5)$								
30	0.437	0.797	0.102	0.155	0.398	0.533	0.599	0.296	0.803
500	0.777	0.826	0.103	0.214	0.420	0.626	0.564	0.082	0.778
2,000	0.799	0.847	0.109	0.237	0.448	0.659	0.551	0.050	0.725

as PTWZ, and to our test as QXZ. In each case, the acronym is derived from the first letters of the last names of the authors. Throughout, we set the sample sizes to be $m = n = 30$, and set the dimension $p = 30$, 90, 150, 200, 500, 1000, 1500, 2000. We repeat our simulations 1,000 times, and report the empirical sizes and powers at the significance level $\alpha = 0.05$.

3.1. Simulation studies

Let $t_d(\mathbf{u}, \boldsymbol{\Sigma})$ represent a multivariate t-distribution with d degrees of freedom, location vector $\mathbf{u}$, and shape matrix $\boldsymbol{\Sigma}$. We draw the p-vectors $\mathbf{x}_i$, for $i = 1, \ldots, m$, from $t_d(\mathbf{u}_1, \boldsymbol{\Sigma}_1)$, and the p-vectors $\mathbf{y}_j$, for $j = 1, \ldots, n$, from $t_d(\mathbf{u}_2, \boldsymbol{\Sigma}_2)$, where $\mathbf{u}_1 = \mathbf{0}_{p\times 1}$, $\boldsymbol{\Sigma}_1 = (0.5^{|k-l|})_{p\times p}$, $\mathbf{u}_2 = \delta \mathbf{1}_{p\times 1}$ and $\boldsymbol{\Sigma}_2 = \sigma^2 \boldsymbol{\Sigma}_1$. We consider four scenarios for (δ, σ^2): $(0, 1)$, $(0.25, 1)$, $(0.15, 2.0)$, and $(0, 2.5)$, which correspond to the null hypothesis H_0 in (1.2), a location shift, both a location shift and a scale difference, and a scale difference only, respectively. For brevity, we report the simulation results for $p = 30$, 500, and 2000 in the main context. The simulation results for $p = 90$, 150, 200, 1000, and 1500 are charted in Tables 1–3, in Appendix G of the Supplementary Material.

Tables 1–3 summarize the empirical sizes and powers for $d = 2$, 3, and 30, respectively, which correspond to heavy, moderate, and light tails, respectively.

Table 2. The empirical sizes and powers when $d = 3$.

	QXZ	PTWZ	R	BMG	BG	BF	HT	H	MBG
p	$(\delta, \sigma^2) = (0, 1.0)$								
30	0.042	0.063	0.055	0.052	0.049	0.050	0.056	0.032	0.052
500	0.031	0.0390	0.070	0.051	0.051	0.044	0.048	0.054	0.056
2,000	0.045	0.054	0.052	0.059	0.046	0.049	0.044	0.040	0.048
p	$(\delta, \sigma^2) = (0.25, 1.0)$								
30	0.228	0.063	0.133	0.101	0.051	0.277	0.060	0.298	0.083
500	0.935	0.043	0.826	0.579	0.053	0.559	0.056	0.820	0.046
2,000	0.992	0.057	0.999	0.958	0.062	0.590	0.053	0.825	0.045
p	$(\delta, \sigma^2) = (0.15, 2.0)$								
30	0.422	0.756	0.105	0.201	0.534	0.652	0.590	0.371	0.781
500	0.943	0.818	0.243	0.453	0.561	0.831	0.544	0.221	0.683
2,000	0.967	0.816	0.483	0.734	0.533	0.863	0.521	0.234	0.632
p	$(\delta, \sigma^2) = (0.0, 2.5)$								
30	0.508	0.921	0.098	0.247	0.750	0.753	0.820	0.368	0.939
500	0.934	0.964	0.113	0.377	0.737	0.907	0.809	0.044	0.886
2,000	0.950	0.967	0.106	0.384	0.747	0.910	0.772	0.022	0.827

Table 3. The empirical sizes and powers when $d = 30$.

	QXZ	PTWZ	R	BMG	BG	BF	HT	H	MBG
p	$(\delta, \sigma^2) = (0, 1.0)$								
30	0.041	0.059	0.065	0.052	0.062	0.053	0.045	0.042	0.048
500	0.057	0.051	0.040	0.045	0.061	0.039	0.049	0.040	0.055
2,000	0.047	0.050	0.057	0.049	0.050	0.060	0.056	0.037	0.060
p	$(\delta, \sigma^2) = (0.25, 1.0)$								
30	0.356	0.071	0.141	0.125	0.054	0.487	0.065	0.311	0.078
500	1.000	0.079	0.840	0.788	0.055	1.000	0.066	0.928	0.067
2,000	1.000	0.071	1.000	0.999	0.073	1.000	0.069	0.965	0.060
p	$(\delta, \sigma^2) = (0.15, 2.0)$								
30	0.628	1.000	0.125	0.444	1.000	0.933	1.000	0.355	1.000
500	1.000	1.000	0.271	0.998	1.000	1.000	1.000	0.000	1.000
2,000	1.000	1.000	0.527	1.000	1.000	1.000	1.000	0.001	0.992
p	$(\delta, \sigma^2) = (0.0, 2.5)$								
30	0.755	1.000	0.121	0.663	1.000	0.991	1.000	0.292	1.000
500	1.000	1.000	0.126	1.000	1.000	1.000	1.000	0.000	1.000
2,000	1.000	1.000	0.153	1.000	1.000	1.000	1.000	0.000	0.994

When $\delta = 0$ and $\sigma^2 = 1$, the sizes of all tests are very close to the significance level $\alpha = 0.05$. The empirical powers of all tests improve quickly as p increases. This is perhaps because the deviations from H_0 in (1.2) are accumulating as p diverges. Furthermore, the R test performs best, followed by the BMG, H, and QXZ tests, when $\delta = 0.25$ and $\sigma^2 = 1$, that is, when only the location shift is present. However, if a scale difference exists, that is, $\sigma^2 = 2$ or 2.5, the R test deteriorates quickly, and our proposed QXZ test exhibits the best power performance. In effect, the PTWZ, BG, HT, and MBG tests are highly insensitive to a location shift in all of our empirical studies, and the H and R tests are highly insensitive to a scale difference. The BG test, which requires the existence of the second moment, performs much better when $d = 30$ than when $d = 2$. In general, our proposed QXZ test exhibits very satisfactory power performance in the presence of a location shift and/or a scale difference.

3.2. Applications

We compare the power performance of the aforementioned tests by analyzing three data sets: a sonar data set, an ECG data set, and a hand outlines data set. The sonar data set is available at `http://www.ics.uci.edu/~mlearn/MLRepository.html`; the other two data sets are available at `http://www.cs.ucr.edu/~eamonn/time_series_data/`. The sonar data contains 111 patterns obtained by bouncing sonar signals off a metal cylinder, together with 97 patterns obtained from rocks. Each number in a 60-dimensional pattern represents the energy within a particular frequency band integrated over a certain period. The ECG data set is a time series recorded at 96 different time points. There are 200 observations, among which 133 are labeled as normal, and the rest as abnormal. The hand outlines data set contains 1,370 observations, each of which is 2,709-dimensional. In this data set, 875 observations are labeled as normal, and 495 are labeled as abnormal.

To compare the power performance of the tests, we randomly select N observations for each class. In other words, we choose $n = m = N$. We consider $N = \{9, 12, 15, 18, 21\}$ as the subsample size for the sonar data set, $N = \{6, 8, 10, 12, 14\}$ for the ECG data set and $N = \{7, 9, 11, 13, 15\}$ for the hand outlines data set. We repeat this random selection procedure 500 times, and report the empirical power of all tests at the significance level $\alpha = 0.05$ in Table 4. Our proposed QXZ test performs much better than its competitors in the sonar data set for $N \geq 9$, and in the ECG data set for $N \geq 6$. In the hand outlines data set, the BF test performs best, closely followed by the PTWZ, H, and proposed QXZ tests. In all applications, the HT and BG tests are the least

Table 4. The powers of all tests for a given size of random samples in the analysis of the sonar dataset, ECG dataset and Hand Outlines dataset.

	QXZ	PTWZ	R	BMG	BG	BF	HT	H	MBG
N	sonar dataset								
9	0.572	0.140	0.232	0.070	0.114	0.230	0.130	0.194	0.162
12	0.720	0.146	0.382	0.194	0.116	0.296	0.130	0.252	0.234
15	0.886	0.216	0.500	0.288	0.156	0.456	0.156	0.452	0.374
18	0.952	0.236	0.658	0.468	0.200	0.548	0.164	0.582	0.494
21	0.980	0.280	0.792	0.526	0.204	0.646	0.216	0.706	0.640
	ECG dataset								
6	0.658	0.612	0.192	0.326	0.624	0.628	0.358	0.558	0.626
8	0.784	0.722	0.154	0.606	0.738	0.740	0.478	0.736	0.768
10	0.892	0.852	0.736	0.538	0.830	0.876	0.592	0.886	0.892
12	0.970	0.938	0.642	0.798	0.930	0.944	0.710	0.936	0.944
14	0.976	0.950	0.930	0.904	0.948	0.960	0.778	0.968	0.966
	Hand Outlines dataset								
7	0.648	0.662	0.420	0.308	0.598	0.702	0.362	0.550	0.594
9	0.808	0.806	0.330	0.466	0.690	0.834	0.584	0.698	0.730
11	0.888	0.930	0.194	0.682	0.834	0.944	0.656	0.834	0.856
13	0.926	0.938	0.590	0.780	0.880	0.954	0.762	0.902	0.912
15	0.960	0.966	0.458	0.686	0.916	0.978	0.836	0.948	0.946

powerful.

4. Conclusion

In this paper, we propose a robust and fully nonparametric two-sample test. Our proposed test is a generalization of the classic Cramér−von Mises test, and can be readily used in high dimensions. It also inherits the advantages of the Cramér−von Mises test in that our proposed test requires no moment condition on the random samples, and is robust to the presence of outliers and extreme values. More importantly, the null distribution of our proposed test statistic is asymptotically standard normal, regardless of the relation between the sample sizes or the dimensions of the two random samples. Therefore, our proposed two-sample test is computationally feasible, in that no bootstrap or re-sampling technique is required to decide critical values in very large-scale two-sample test problems.

We use the Cramér−von Mises distance rather than the Kolmogorov−Smirnov distance. This accommodates the U-statistic theory, and allows us to study the asymptotic behavior of the test statistic systematically. However, generalizing our idea of handling high-dimensional two-sample test problems to use the Kolmogorov−Smirnov distance is also important, and thus warrants further research.

Supplementary Material

The proofs of (2.11)−(2.13), and Theorem 1–Theorem 3 and additional simulation results are provided in the online Supplement Material.

Acknowledgments

This work was supported by the Beijing Natural Science Foundation (Z19J00 009), National Natural Science Foundation of China (11731011, 11931014, and 11971478), and MOE Project of the Key Research Institute of Humanities and Social Sciences at Universities (16JJD910002). All authors contributed equally to this paper. Correspondence should be addressed to Liping Zhu.

References

Anderson, T. W. (1962). On the distribution of the two-sample cramér-von mises criterion. *Annals of Mathematical Statistics* **34**, 1148–1159.

Baringhaus, L. and Franz, C. (2004). On a new multivariate two-sample test. *Journal of Multivariate Analysis* **88**, 190–206.

Biswas, M. and Ghosh, A. (2014). A nonparametric two-sample test applicable to high dimensional data. *Journal of Multivariate Analysis* **123**, 160–171.

Biswas, M., Mukhopadhyay, M. and Ghosh, A. (2014). A distribution-free two-sample run test applicable to high-dimensional data. *Biometrika* **101**, 913–926.

Chen, H. and Friedman, J. (2017). A new graph-based two-sample test for multivariate and object data. *Journal of the American Statistical Association* **112**, 397–402.

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

Friedman, J. and Rafsky, L. (1979). Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests. *The Annals of Statistics* **7**, 697–717.

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B. and Smola, A. (2012). A kernel two-sample test. *Journal of Machine Learning Research* **13**, 723–773.

Hall, P. and Heyde, C. (1980). *Martingale Limit Theory and Its Applications*. 1st Edition. Academic Press, New York.

Hall, P. and Tajvidi, N. (2002). Permutation tests for equality of distributions in high-dimensional settings. *Biometrika* **89**, 359–374.

Henze, N. (1988). A multivariate two-sample test based on the number of nearest neighbor type coincidences. *The Annals of Statistics* **16**, 772–783.

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

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypothesis*. 3rd Edition. Springer, New York.

Mondal, P., Biswas, M. and Ghosh, A. (2015). On high dimensional two-sample tests based on nearest neighbors. *Journal of Multivariate Analysis* **141**, 168–178.

Pan, W., Tian, Y., Wang, X. and Zhang, H. (2018). Ball divergence: Nonparametric two sample test. *The Annals of Statistics* **46**, 1109–1137.

Rosenbaum, P. (2005). An exact distribution-free test comparing two multivariate distributions based on adjacency. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **67**, 515–530.

Rosenblatt, M. (1952). Limit theorems associated with variants of the von mises statistic. *Annals of Mathematical Statistics* **23**, 617–623.

Schilling, M. F. (1986). Multivariate two-sample tests based on nearest neighbors. *Journal of the American Statistical Association* **81**, 799–806.

Smirnov, N.V. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Moscow University Mathematics Bulletin* **2**, 3–14.

Thas, O. (2010). *Comparing Distributions.* Springer, New York.

Zhang, X., Yao, S. and Shao, X. (2018). Conditional mean and quantile dependence testing in high dimension. *The Annals of Statistics* **46**, 219–246.

Zhong, P. and Chen, S. (2011). Tests for high-dimensional regression coefficients with factorial designs. *Journal of the American Statistical Association* **106**, 260–274.

Tao Qiu

School of Statistics, Renmin University of China, Beijing 100872, China.

E-mail: qiutao8505@126.com

Wangli Xu

School of Statistics, Renmin University of China, Beijing 100872, China.

E-mail: wlxu@ruc.edu.cn

Liping Zhu

Center for Applied Statistics, Institute of Statistics and Big Data, Renmin University of China, Beijing 100872, China.

School of Statistics and Mathematics, Zhejiang Gongshang Uinversity, Hangzhou 310018, China.

E-mail: zhu.liping@ruc.edu.cn

(Received October 2017; accepted April 2019)