A rank-based Cramér–von-Mises-type test for two samples

Jamye Curry[a], Xin Dang[b] and Hailin Sang[b]

[a]*Georgia Gwinnett College*
[b]*University of Mississippi*

Abstract. We study a rank based univariate two-sample distribution-free test. The test statistic is the difference between the average of between-group rank distances and the average of within-group rank distances. This test statistic is closely related to the two-sample Cramér–von Mises criterion. They are different empirical versions of a same quantity for testing the equality of two population distributions. Although they may be different for finite samples, they share the same expected value, variance and asymptotic properties. The advantage of the new rank based test over the classical one is its ease to generalize to the multivariate case. Rather than using the empirical process approach, we provide a different easier proof, bringing in a different perspective and insight. In particular, we apply the Hájek projection and orthogonal decomposition technique in deriving the asymptotics of the proposed rank based statistic. A numerical study compares power performance of the rank formulation test with other commonly-used nonparametric tests and recommendations on those tests are provided. Lastly, we propose a multivariate extension of the test based on the spatial rank.

1 Introduction

To test whether two samples come from the same or different populations, several distribution free tests such as the Kolmogorov–Smirnov test, the Cramér–von Mises test and their variations have been proposed and widely used. Let $X_1, X_2, \ldots, X_m \overset{\text{i.i.d.}}{\sim} F$ and $Y_1, Y_2, \ldots, Y_n \overset{\text{i.i.d.}}{\sim} G$ be two independent random samples with continuous distribution functions F and G, respectively. The two sample problem is to test

$$H_0 : F = G \quad \text{vs} \quad H_a : F \neq G. \tag{1}$$

Denote F_m and G_n as the empirical distribution functions of the two samples and H_N as the empirical distribution function of the combined sample, where $N = m + n$. The Kolmogorov–Smirnov (KS) two-sample test uses the maximum distance (difference) between F_m and G_n. The classical Cramér–von Mises test statistic has the form

$$T_c = \frac{mn}{N} \int_{-\infty}^{\infty} \left[F_m(x) - G_n(x) \right]^2 dH_N(x). \tag{2}$$

This test statistic and its asymptotics have been well studied in the literature, for example, Lehmann (1951), Rosenblatt (1952), Darling (1957), Fisz (1960) and Anderson (1962).

Both of the KS test statistic and the Cramér–von Mises test statistic are formulated based on the empirical distributions. Székely and Rizzo (2004), Baringhaus and Franz (2004) studied a test statistic based on the original data. That is

$$\frac{mn}{N} \left\{ \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} |X_i - Y_j| - \frac{1}{2m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} |X_i - X_j| \right.$$
$$\left. - \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} |Y_i - Y_j| \right\}. \tag{3}$$

This test has a direct generalization to the multivariate case. However, it requires an assumption on the first moment and it is not distribution free for the univariate case. It is worth to note that the test statistic (3) falls in the unified framework on energy statistics studied by Székely and Rizzo (2013, 2017) and can be easily generalized to the K sample problem. Other similar tests include Fernández, Jimènez Gamerro and Muñoz Garcìa (2008) and Gretton et al. (2008), although they are derived under different motivations. Fernández, Jimènez Gamerro and Muñoz Garcìa (2008) developed a statistic based on the empirical characteristic functions of the observed observations. The statistic uses a weighted integral of the difference between the empirical characteristic function of the two samples. Gretton et al. (2008) proposed a test based on a kernel method in which the testing procedure is defined as the maximum difference in expectations over functions evaluated on the two samples. All of those test statistics are of the form being a difference on a measure of between-group and within-group.

In this paper, we propose a new rank based test of the same form. Nevertheless, it overcomes the limitations of (3). It is formulated based on the ranks of two samples with respect to the combined sample H_N. Denote $R(y, H)$ as the standardized rank of the quantity y with respect to the distribution H, that is, $R(y, H) = H(y)$. For testing the hypothesis (1), we use the following test statistic.

$$T = \frac{mn}{N} \left\{ \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \left| R(X_i, H_N) - R(Y_j, H_N) \right| \right.$$
$$- \frac{1}{2m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} \left| R(X_i, H_N) - R(X_j, H_N) \right| \tag{4}$$
$$\left. - \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left| R(Y_i, H_N) - R(Y_j, H_N) \right| \right\}.$$

T is interpreted as the difference of the average of between-group rank differences and the average of within-group rank differences. A large value of T indicates

the deviation of two groups. The test based on T is distribution-free and does not require any moment condition.

For the balanced samples ($m = n$), one can consider an equivalent but simpler statistic

$$T' = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \left| R(X_i, H_N) - R(Y_j, H_N) \right|. \tag{5}$$

T' is the average of rank differences between two groups. T and T' are equivalent because $T = nT' - (4n^2 - 1)/(12n)$ when $m = n$.

As we will see later, the test statistic T is closely related to the classical non-parametric Cramér–von Mises criterion T_c. They are different empirical plug-in versions of the same population quantity. The rank based test statistic and the Cramér–von Mises criterion may not be equal to each other for finite samples, but they are asymptotically equivalent. The advantage of the new rank based test over the classical one is its ease to generalize to the multivariate case. Multivariate generalizations of Cramér–von Mises tests have been considered by many researchers, but they are either applied on independent data (Cotterill and Csörgő (1982)) or used for testing independence (Genest, Quessy and Rémillard (2007)) or used in the goodness-of-fit test of the uniform distribution on the transformed data (Chiu and Liu (2009)). For the rank based formulation, generalizations to the multivariate two sample problem are straightforward by applying notions of multivariate rank functions. In this paper, rather than using the empirical process approach, we provide a different easier proof, bringing in a different perspective and insight. In particular, we apply the Hájek projection and orthogonal decomposition technique in deriving the asymptotics of the proposed statistic.

Some related works include Pettitt (1976) and Baumgartner, Weiß and Schindler (1998). They considered statistics of Anderson–Darling type that can be viewed as standardized versions of Cramér–von Mises statistics. Schmid and Trede (1995) utilized $\mathcal{L}^1$ Cramér–von Mises statistics. A rank-based representation of a $\mathcal{L}^1$ Cramér–von Mises statistic under a balanced size and its generalizations are studied by Borroni (2001). Albers, Kallenberg and Martini (2001) studied rank procedures for detecting shift alternatives with increasing shift in the tail of the distribution. Janic-Wróblewska and Ledwina (2000) considered a test based on a combination of several linear rank statistics. Related to the rank procedures, other nonparametric tests include those based on the empirical likelihood approach. Einmahl and McKeague (2003) considered test statistics based on the empirical likelihood ratios for the goodness of fit and two sample problems. It has been proved that those tests are asymptotically equivalent to the one-sample and two sample Anderson–Darling tests. Cao and Van Keilegom (2006) proposed an empirical likelihood ratio test via kernel density estimation. Gurevich and Vexler (2011) utilized an empirical likelihood ratio test based on samples entropy.

The paper has the following structure. Section 2 presents the main results, including the formulation of the test statistic and its properties. The simulation study

is performed in Section 3. We propose a multivariate extension of the test in Section 4. We summarize and conclude the paper in Section 5. All proofs go to Section 6.

2 Main results

To formulate the rank based test statistic T in (4), we first establish its population version. We provide a result of the population version, from which we can see the relationship between our statistic and Cramér–von Mises criterion.

Theorem 2.1. *Let X, X_1, X_2 and Y, Y_1, Y_2 be independent continuous random variables distributed from F and G, respectively. Let $H = \tau F + (1 - \tau)G$ with $0 \le \tau \le 1$ be the mixture distribution. Then*

$$\mathbb{E}\big|R(X, H) - R(Y, H)\big| - \frac{1}{2}\mathbb{E}\big|R(X_1, H) - R(X_2, H)\big|$$
$$- \frac{1}{2}\mathbb{E}\big|R(Y_1, H) - R(Y_2, H)\big| \ge 0 \tag{6}$$
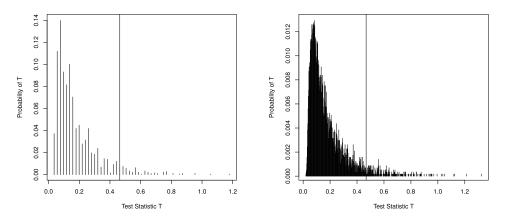
and the equality holds if and only if $F = G$.

The above result is based on the following identity which is obtained from Lemma 6.1 in the Proofs section.

$$\mathbb{E}\big|R(X, H) - R(Y, H)\big| - \frac{1}{2}\mathbb{E}\big|R(X_1, H) - R(X_2, H)\big|$$
$$- \frac{1}{2}\mathbb{E}\big|R(Y_1, H) - R(Y_2, H)\big| \tag{7}$$
$$= \int_{-\infty}^{\infty} \big(F(x) - G(x)\big)^2 \, d\big(\tau F(x) + (1 - \tau)G(x)\big).$$

The result of Theorem 2.1 suggests two possible statistics for testing the hypothesis (1). The first version is the sample plug-in version of the left side of (7). With $\tau = m/N$ and multiplying by mn/N, it is our test statistic defined in (4). H_0 is rejected if the sample version is large, i.e., $T > c_\alpha(m, n)$. The critical value $c_\alpha(m, n)$ is determined by the significance level α and the null distribution of T. The test statistic T is the difference of the average of between-group rank differences and the average of within-group rank differences. A large value of T indicates the deviation of two groups.

The two-sample Cramér–von Mises statistic T_c in (2) is the empirical version of the right-hand side of (7). Hence, T and T_c are all plug-in statistics of an equal quantity. Nevertheless, they may take different values. We shall thank one of the referees who pointed out this possibility. For example, in the case that $m = n = 2$, let the two X realizations be 0 and 2 and the two Y realizations be 1 and 3. It is

Figure 1 *The exact null distribution of T obtained from all combinations (Left: m = 7, n = 7; Right: m = 7, n = 9). The vertical line in each graph indicates the 5% critical value.*

easy to see that the Cramér–von Mises statistic has value $\frac{1}{4}$ and the test statistic T has value $\frac{1}{8}$. Next, we will study the properties of T.

Let D be $\mathbb{E}|R(X, H) - R(Y, H)| - \frac{1}{2}\mathbb{E}|R(X_1, H) - R(X_2, H)| - \frac{1}{2}\mathbb{E}|R(Y_1, H) - R(Y_2, H)|$, and $\hat{D} = N/(mn)T$. Then we have the following theorem.

Theorem 2.2. *For $m, n \to \infty$, if $m/(m + n) \to \tau$, then $\hat{D} \to D$ a.s.*

By this theorem and Theorem 2.1, it is easy to see that our test statistic T is consistent for the alternative $H_a : F \neq G$.

Theorem 2.3. *Under H_0, T is distribution free.*

Under H_0, the combined samples $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ constitute a random sample of size N from the distribution $F = G = H$. So any assignment of m numbers to $X_1, \ldots, X_m$ and n numbers to $Y_1, \ldots, Y_n$ from the set of integers $\{1, 2, \ldots, N\}$ is equally likely, that is, has probability $\binom{N}{m}^{-1}$ which is independent of F. Using the fact that those number assignments have one-to-one linear relationship with the standardized ranks, T is distribution free.

The exact null distribution of T can be found by enumeration of all possible values of T by considering the $N!/(m!n!)$ orderings of m X's and n Y's. Figure 1 provides the exact null distribution of T for sample sizes $m = n = 7$ and $m = 7, n = 9$ by considering all combinations. However, the exact null distribution is infeasible to obtain for large sample sizes because the number of combinations increases dramatically as m and n increase. For large samples, we can use Monte-Carlo method on all combinations to approximate the null distribution. Also the limiting distribution of T can be used to determine the critical values of the test. Next, we study asymptotic behaviors of T.

Since T and the Cramér–von Mises statistic T_c are different sample plug-in versions of a same quantity for checking the equality of two distributions, we expect that they should have the same expectation, variance and asymptotic distribution under the null hypothesis. The expectation and variance of T_c are provided by Anderson (1962). In the next theorem, we obtain the same results for T, but provide a more straightforward derivation and simpler proof.

Theorem 2.4. *Under H_0,*

$$\mathbb{E}T = \frac{N+1}{6N} = \frac{1}{6} + \frac{1}{6N}$$

and

$$\mathrm{Var}(T) = \frac{N+1}{180N^2}\left[4(N+1) - \frac{3N^2}{mn}\right].$$

Remark 2.1. In particular, if $m, n \to \infty$, $\mathbb{E}T \to 1/6$ and $\mathrm{Var}(T) \to 1/45$.

Rosenblatt (1952) and Fisz (1960) have derived the limiting distribution of Cramér–von Mises statistic, which is a mixture of independent χ_1^2 distributions. It is necessary to check whether or not our test statistic has the same limiting distribution. Rather than using a stochastic process method, we provide a different Hájek projection approach to obtain the limiting distribution of T, which agrees with that of T_c.

We obtain the first order Hájek projection $\tilde{T}$ in Lemma 6.2 as

$$\tilde{T} = \sum_{i=1}^{m} \mathbb{E}[T|X_i] + \sum_{j=1}^{n} \mathbb{E}[T|Y_j] - (N-1)\mathbb{E}T,$$

where $\mathbb{E}[T|X_1] = \frac{1}{6}(1 + \frac{n}{mN}) + \frac{1}{N}(1 - \frac{n}{m})[F(X_1) - F^2(X_1)]$. Under H_0, $\tilde{T}$ has variance

$$\begin{aligned}
\mathrm{Var}(\tilde{T}) &= \sum_{i=1}^{m} \mathrm{Var}\big(\mathbb{E}[T|X_i]\big) + \sum_{j=1}^{n} \mathrm{Var}\big(\mathbb{E}[T|Y_j]\big) \\
&= \left[\frac{m(m-n)^2}{m^2N^2} + \frac{n(m-n)^2}{n^2N^2}\right]\mathrm{Var}\big[F(Y_1) - F^2(Y_1)\big] \\
&= \frac{(m-n)^2}{180mnN}.
\end{aligned}$$

Clearly, $\mathrm{Var}(\tilde{T})/\mathrm{Var}(T) \to 0$ as $m, n \to \infty$. Therefore, the first order Hájek projection is not sufficient in deriving the asymptotics of the statistic T.

To derive the asymptotics of the statistic T under the null hypothesis, it is necessary to have the second order projection $\hat{T}$ of T.

$$\hat{T} = \sum_{i=1}^{m} \sum_{j=1}^{n} \mathbb{E}[T|X_i, Y_j] + \sum_{1 \leq i < j \leq m} \mathbb{E}[T|X_i, X_j]$$

$$+ \sum_{1 \leq i < j \leq n} \mathbb{E}[T|Y_i, Y_j] - \frac{N(N-1)}{2} \mathbb{E}T.$$

Since

$$\mathbb{E}\{\mathbb{E}[T|X_1, Y_1]\} = \mathbb{E}\{\mathbb{E}[T|X_1, X_2]\} = \mathbb{E}\{\mathbb{E}[T|Y_1, Y_2]\} = \mathbb{E}T,$$

$\mathbb{E}\hat{T} = 0$. By Lemma 6.3 and Lemma 6.4, it can be examined that

$$\text{Cov}\big(\mathbb{E}[T|Z_1, Z_2], \mathbb{E}[T|Z_1, Z_3]\big) = 0,$$

where Z_1, Z_2 and Z_3 are three different variables from $X_i, Y_j, 1 \leq i \leq m, 1 \leq j \leq n$, and

$$\text{Var}(\hat{T}) = \frac{m^4 + n^4 - 2m^3 n - 2mn^3 + 10m^2 n^2 - 8mnN + 5n^2 + 5m^2}{180 N^2 mn}$$

$$+ \frac{m(m-1)}{2} \frac{m^2 - 2mn + 5n^2}{90 m^2 N^2} + \frac{n(n-1)}{2} \frac{n^2 - 2mn + 5m^2}{90 n^2 N^2}$$

$$= \frac{N^2}{180 mn} + \frac{2}{45 N} - \frac{N-1}{36 mn} - \frac{1}{18 N^2}.$$

Then $\frac{\text{Var}(\hat{T})}{\text{Var}(T)} \to 1$ as $N \to \infty$ under the condition $\lim_{N \to \infty} m/n = 1$. We shall always assume this condition in the following analysis.

Efron and Stein (1978) discussed a general orthogonal decomposition of a statistic. Here, our statistic T is decomposed as $\tilde{T} + \hat{T} + R_N$, where $\tilde{T}$ is the first order projection and R_N is a negligible term. Hence, the limiting distribution of T is determined by the limiting distribution of $\hat{T}$.

To determine the limiting distribution of $\hat{T}$ under H_0, let $h(x, y) = |F(x) - F(y)| + F(x)[1 - F(x)] + F(y)[1 - F(y)] - 2/3$. Then $h(x, y)$ is a degenerate kernel function since $h(x, y)$ is symmetric and $\mathbb{E}h(X, y) = 0$. By Lemma 6.3 and Lemma 6.4, we have $\hat{T} = \hat{\tilde{T}} + R'_N$ with $\text{Var}(R'_N)/\text{Var}(T) \to 0$ as $N \to \infty$[1] and

$$\hat{\tilde{T}} = \frac{1}{N} \sum_{i=1}^{m} \sum_{j=1}^{n} h(X_i, Y_j) - \frac{1}{N} \sum_{1 \leq i < j \leq m} h(X_i, X_j) - \frac{1}{N} \sum_{1 \leq i < j \leq n} h(Y_i, Y_j).$$

It is not difficult to verify that

$$\text{Var}\big[h(Z_1, Z_2)\big] = 2/45 \tag{8}$$

---

[1] If $m = n \to \infty$, $R'_N = 0$.

and $\mathrm{Cov}(h(Z_1, Z_2), h(Z_1, Z_3)) = 0$, that is, $h(Z_1, Z_2)$ and $h(Z_1, Z_3)$ are orthogonal, where Z_1, Z_2 and Z_3 are three different variables from X_i, Y_j, $1 \le i \le m$, $1 \le j \le n$.

Now we define an operator A on the function space $\mathcal{L}^2(\mathbb{R}, F)$ by

$$Ag^*(x) = \int_{-\infty}^{\infty} h(x, y)g^*(y)\, dF(y), \qquad x \in \mathbb{R}, g^* \in \mathcal{L}^2(\mathbb{R}, F).$$

This operator only has real eigenvalues since the kernel $h(x, y)$ is symmetric. Let $\lambda = \lambda_1, \lambda_2, \ldots$ be the nonzero eigenvalues of the operator A obtained by solving the equation $Ag^* = \lambda g^*$. With the substitution of $u = F(x)$ and $v = F(y)$, solving $Ag^* = \lambda g^*$ is equivalent to solve that

$$\int_0^1 \left\{ |u - v| + u(1 - u) + v(1 - v) - \frac{2}{3} \right\} g(v)\, dv = \lambda g(u), \tag{9}$$

where $g = g^* \circ F^{-1}$. Taking the twice derivative with respect to u on both sides of (9), we have the equation $2g(u) = \lambda g''(u)$. Solving it and substituting back, we have the eigenvalues of A being $\lambda_k = -\frac{2}{\pi^2 k^2}$, $k \in \mathbb{N}$ and the corresponding eigenfunctions $\phi_k(x) = \cos(k\pi F(x))$, $k \in \mathbb{N}$. The eigenfunction for the zero eigenvalue is $\phi_0(x) = 1$. Note that the eigenvalues do not depend on F, but the eigenfunctions $\{\phi_k(x)\}_{k=0}^{\infty}$ depend on F, which give a orthonormal basis for the space $\mathcal{L}^2(\mathbb{R}, F)$. Let $T_N = \hat{\hat{T}}/\sqrt{\mathrm{Var}(T)}$. Then we have the following theorem.

Theorem 2.5. *Under H_0 and the condition $\lim_{N \to \infty} m/n = 1$,*

$$T_N \xrightarrow{d} Z_\infty = -\frac{\sqrt{45}}{2} \sum_{k=1}^{\infty} \lambda_k (\chi_{1k}^2 - 1),$$

where $\chi_{11}^2, \chi_{12}^2, \ldots$ are independent χ_1^2 variables and $\lambda_k = -\frac{2}{\pi^2 k^2}$, $k \in \mathbb{N}$. Hence

$$(T - \mathbb{E}T)/\sqrt{\mathrm{Var}(T)} \xrightarrow{d} Z_\infty = -\frac{\sqrt{45}}{2} \sum_{k=1}^{\infty} \lambda_k (\chi_{1k}^2 - 1)$$

since $(T - \mathbb{E}T - \hat{\hat{T}})/\sqrt{\mathrm{Var}(T)} \to 0$ in probability.

As expected, this asymptotical result agrees with the one for Cramér–von Mises statistic as proved with a stochastic process method in Rosenblatt (1952) and Fisz (1960). This different projection approach we applied here is typically useful in U-statistic theory, but we shall emphasize that T is not an U-statistic.

In practice, we may approximate the limiting distribution by a distribution of a finite linear combination of d independent χ_1^2 random variables, that is,

$$Z_d = -\frac{\sqrt{45}}{2} \sum_{k=1}^{d} \lambda_k (\chi_{1k}^2 - 1) = \frac{\sqrt{45}}{\pi^2} \sum_{k=1}^{d} \frac{1}{k^2} (\chi_{1k}^2 - 1).$$

Table 1 *First part*: variance ratios of Z_d over Z_∞ and 95% quantiles of Z_d. Second part: approximated critical values for T. Comparing with the exact $\alpha = 0.049$ critical value 0.4643 for the case of $m = n = 7$ and the exact $\alpha = 0.05$ critical value 0.4678 for the case of $m = 7, n = 9$, the approximations are pretty accurate even under small sizes. In practice, $d = 4$ or $d = 10$ is recommended

	$d = 1$	$d = 2$	$d = 4$	$d = 10$	$d = 100$
Variance ratio	0.9239	0.9819	0.9967	0.9997	1.0000
95% quantile	1.9298	1.9676	1.9772	1.9779	1.9780
Approximated $c_\alpha(m, n)$ of T ($\alpha = 0.05$) based on Z_d					
$m = n = 50$	0.4545	0.4601	0.4617	0.4617	0.4617
$m = 50, n = 40$	0.4545	0.4601	0.4615	0.4616	0.4616
$m = n = 500$	0.4544	0.4600	0.4614	0.4615	0.4615
$m = n = 7$	0.4543	0.4597	0.4610	0.4611	0.4611
$m = 7, n = 9$	0.4540	0.4594	0.4608	0.4609	0.4609

The accuracy of approximation depends on the choice of d. Table 1 provides ratios of variance of the d mixture and that of the infinite mixture, that is, $\sigma^2(Z_d)/\sigma^2(Z_\infty)$. Also the table lists 95% quantiles of Z_d which are estimated by the average of 10 sample quantiles each on $M = 10^8$ random samples. Those quantile values can be used to approximate the critical values $c_\alpha(m, n)$ of T, which are given by the second part of Table 1. As we will see that even for small sample sizes, the approximated critical values are pretty accurate and close to the exact true values. For the case of $m = n = 7$, the true size of the test is 0.056 if the approximated critical value 0.4611 is used. For the case of $m = 7$ and $n = 9$, the true size of the test is 0.052 if 0.4609 is used. In summary, $d = 4$ or $d = 10$ is recommended for a compromise between computation and accuracy.

3 Simulations

By the simulation study in this section, we demonstrate the performance of the T test. There are many nonparametric tests available for the two sample problem. It is by no means to conduct a comprehensive comparison. Here we include Kolmogorov–Smirnov test (KS), Wilcoxon rank sum test (W) or Mood test (M), the empirical likelihood ratio test (ELR) proposed by Gurevich and Vexler (2011), the empirical likelihood test (ELT) proposed by Einmahl and McKeague (2003), Baringhaus and Franz's Cramér test (CT), the test studied in Fernández, Jimènez Gamerro and Muñoz Garcìa (2008) (DT) in the study. It is necessary to note that the CT and DT tests are not distribution-free tests, and their critical values and p-values are based Monte-Carlo method on permutations in each sample, which is implemented in the R package "cramer". The R package "dbEmpLikeGOF" is used for the ELR test in which the parameter is set to be 0.1 as suggested in Gurevich and Vexler (2011). The critical values of the ELT and our T test are computed through 10^7 random combinations on $\{1, \ldots, N\}$.

Table 2 *Power performance of each test with significance level $\alpha = 0.05$ for the normal distribution with location alternatives. Row* 1: $n = m = 50$, *Row* 2: $n = 50$, $m = 40$

Δ	KS	W	ELR	ELT	CT	DT	T
0	0.040	0.050	0.057	0.050	0.050	0.051	0.050
	0.041	0.047	0.031	0.047	0.048	0.048	0.049
0.25	0.162	0.228	0.182	0.224	0.226	0.191	0.217
	0.160	0.208	0.119	0.196	0.203	0.171	0.198
0.5	0.534	0.681	0.578	0.670	0.671	0.582	0.652
	0.498	0.621	0.446	0.603	0.615	0.526	0.600
0.75	0.875	0.949	0.902	0.945	0.943	0.901	0.936
	0.851	0.926	0.829	0.919	0.922	0.871	0.912
1	0.988	0.998	0.994	0.997	0.998	0.991	0.996
	0.979	0.995	0.976	0.994	0.994	0.984	0.993

Various alternative distributions are considered. For each case, $M = 10{,}000$ iterations are computed to estimate powers by calculating the fraction of p-values less than or equal to $\alpha = 0.05$ the level of significance. The Monte Carlo errors can be estimated by $\pm 1.96\sqrt{p(1-p)/M}$. In particular, the size of tests shall maintain in the interval $(0.046, 0.054)$. Table 2 shows the size and power performance for each test under the normal distributions, where $X_1, \ldots, X_n \sim N(0, 1)$ and $Y_1, \ldots, Y_m \sim N(\Delta, 1)$ with $\Delta = 0, 0.25, 0.5, 0.75$, and 1. When $\Delta = 0$, the KS test is undersized for both the equal and unequal sample sizes cases; the ELR test is oversized in the equal sample size case and seriously undersized for the sample unequal size case; all other tests keep a desirable size. As expected, the W test is the best among all tests since it is well-known to be powerful for the two-sample problem with a constant shift in location, especially when data follow logistic or normal distributions. The CT and ELT tests are comparable to W. The T test is more powerful than the DT, KS and ELR tests. In the unequal sample size case, the W test is the best followed by the CT test. The ELT and T tests are comparable and significantly better than the DT, KS and ELR tests.

The experiment is repeated for the t-distribution with 3 degrees of freedom and the result is presented in Table 3. Although the statistical power of the T test is the highest among all tests for all cases, its power differences with the W test or the ELT test are small so that those three tests are comparable.

Table 4 shows the power performance for the Pareto distribution, where $X_1, \ldots, X_n \sim \mathrm{Pa}(2, 2)$ and $Y_1, \ldots, Y_m \sim \mathrm{Pa}(2 + \Delta, 2)$ are generated, with $\Delta = 0, 0.25, 0.5, 0.75$, and 1. The power of the ELR test is much higher than that of all others. For $\Delta = 0.25$, the power of the ELR test is as high as 90%, which is 30% higher than the second best ELT test. The T test is the third best one. The power difference between the T test and that of the CT test can be as large as 27% for equal sample sizes and can be as large as 32% for unequal sample sizes.

Table 3 *Power performance of each test with significance level $\alpha = 0.05$ for the t_3 with location alternatives. Row* 1: $n = m = 50$, *Row* 2: $n = 50, m = 40$

Δ	KS	W	ELR	ELT	CT	DT	T
0	0.036	0.048	0.052	0.048	0.049	0.046	0.047
	0.045	0.054	0.036	0.054	0.055	0.051	0.054
0.25	0.130	0.163	0.124	0.160	0.158	0.142	0.165
	0.135	0.157	0.084	0.152	0.154	0.137	0.162
0.5	0.422	0.488	0.367	0.486	0.481	0.434	0.501
	0.402	0.449	0.268	0.440	0.439	0.394	0.462
0.75	0.776	0.830	0.710	0.825	0.818	0.783	0.836
	0.741	0.786	0.586	0.780	0.776	0.734	0.799
1	0.947	0.966	0.917	0.966	0.965	0.950	0.971
	0.929	0.946	0.842	0.946	0.944	0.925	0.954

Table 4 *Power performance of each test with significance level $\alpha = 0.05$ for the Pareto distributions with location alternatives. Row* 1: $n = m = 50$, *Row* 2: $n = 50, m = 40$

Δ	KS	W	ELR	ELT	CT	DT	T
0	0.040	0.052	0.058	0.051	0.052	0.050	0.051
	0.044	0.050	0.032	0.050	0.050	0.053	0.052
0.25	0.417	0.443	0.906	0.599	0.205	0.287	0.472
	0.377	0.405	0.815	0.516	0.186	0.265	0.428
0.5	0.960	0.886	0.999	0.978	0.655	0.843	0.968
	0.940	0.859	0.998	0.958	0.598	0.798	0.948
0.75	0.999	0.989	1.000	0.999	0.945	0.993	0.999
	0.998	0.980	1.000	0.998	0.908	0.988	0.997
1	1.000	0.999	1.000	1.000	0.993	1.000	1.000
	1.000	0.998	1.000	1.000	0.988	1.000	1.000

All considered tests as in the experiment for location alternatives are used for scale alternatives except the Wilcoxon test (W), as this is a test for location. Instead, the Mood's test known as a scale test is used and referred to as the M test. Table 5 displays the results when Y samples of size 50 are generated from $N(0, \Delta)$ or Pareto$(2, 2\Delta)$, where $\Delta = 1, 1.5, 2, 2.5$, and 3. In the normal case, the T test does not compare favorably to all considered tests other than the KS test. It performs significantly better than the KS test, but its power is 2–5 times smaller than that of others. It is interesting to see that the M test outperforms all tests in the normal case but it is the inferior in the Pareto case. The T test has better performance for Pareto samples than for normal samples due to the heavy tails in Pareto distributions. In the Pareto case, all tests outperform the M test by a great margin and the CT test is the superior. As suggested by a reviewer, we add one more case in the simulation in which $X_1, \ldots, X_n \sim \exp(1)$ and $Y_1, \ldots, Y_m \sim \text{lognorm}(0, 1)$ with sample sizes $m = n = 50$. The Monte Carlo powers of the seven tests are

Table 5 *Power performance of each test with significance level* $\alpha = 0.05$ *for Normal and Pareto scale alternatives*, *also the case of F* = Exp *and G* = Lognorm

Distribution	Δ	KS	M	ELR	ELT	CT	DT	T
Normal	1	0.039	0.051	0.056	0.047	0.049	0.051	0.047
	1.5	0.118	0.663	0.542	0.238	0.251	0.431	0.138
	2	0.374	0.979	0.965	0.746	0.792	0.915	0.479
	2.5	0.681	0.999	0.999	0.962	0.981	0.994	0.815
	3	0.881	1.000	1.000	0.996	0.999	1.000	0.957
Pareto	1	0.040	0.054	0.055	0.049	0.052	0.049	0.050
	1.5	0.307	0.098	0.378	0.418	0.487	0.356	0.398
	2	0.741	0.165	0.828	0.857	0.909	0.815	0.831
	2.5	0.937	0.214	0.973	0.980	0.992	0.974	0.978
	3	0.988	0.234	0.997	0.998	0.999	0.997	0.998
Exp vs Lognorm		0.336	0.535	0.654	0.555	0.502	0.315	0.476

listed in Table 5. In this scenario, the T test performs better than KS and DT, but does not compare as favorably to the CT, W, ELT and ELR tests.

In general, the T test is not recommended for scale alternatives. The Kolmogorov–Smirov test is not recommended either. The empirical likelihood ELR test is more suitable for a general scale alternative, but is not recommended for a location alternative for symmetric distributions. The T test has a better performance for location alternatives than scale alternatives. It is easy to explain the power performance of the Cramér–von Mises test with the rank based formulation (4) for the location alternatives. For two samples from the same class distributions (normal distributions, t distributions or Pareto distributions and so on) but with different locations, the ranks in the mixture are quite different. Therefore the corresponding test can easily recognize them and have good power performance. We recommend to apply the T test for location alternatives, especially in the heavy-tailed distributions.

4 Multivariate extension

The proposed rank test statistic is closely related to the two sample Cramér–von Mises criterion. Both statistics are different sample plug-in forms from a same population quantity. The advantage of our rank test is to allow straightforward generalizations to the multivariate case by using different multivariate rank functions. Among them, the spatial rank is appealing due to its computation ease, efficiency and other nice properties (Möttönen, Oja and Tienari (1997), Oja (2010)). The sample version of the spatial rank function with respect to H_N, the empirical distribution of the combined sample $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$ in $\mathbb{R}^d$, is defined

as

$$R(x, H_N) = \frac{1}{N} \sum_{i=1}^{N} \frac{x - z_i}{\|x - z_i\|},$$

where $z_i = x_i$ for $i = 1, \ldots, m$, $z_{m+i} = y_i$ for $i = 1, \ldots, n$ and $\|\cdot\|$ is the Euclidian distance. Then the multivariate two-sample spatial rank statistic, denoted by T_M, is defined as

$$
\begin{aligned}
T_M = \frac{mn}{N} \Bigg\{ &\frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \|R(x_i, H_N) - R(y_j, H_N)\| \\
&- \frac{1}{2m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} \|R(x_i, H_N) - R(x_j, H_N)\| \\
&- \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \|R(y_i, H_N) - R(y_j, H_N)\| \Bigg\}.
\end{aligned}
\tag{10}
$$

The test statistic T_M is the difference of the average of the intra-group rank distances and the average of the inter-group rank distances. A large value of T_M indicates the deviation of the two groups and rejects the null hypothesis. The multivariate counterpart of Theorem 2.1 states as follows.

Theorem 4.1. *Let X, X_1, X_2 and Y, Y_1, Y_2 be independent d-variate continuous random vectors distributed from F and G, respectively. Let $H = \tau F + (1 - \tau)G$ with $0 \leq \tau \leq 1$. Then*

$$
\begin{aligned}
\mathbb{E}\|R(X, H) - R(Y, H)\| &- \frac{1}{2}\mathbb{E}\|R(X_1, H) - R(X_2, H)\| \\
&- \frac{1}{2}\mathbb{E}\|R(Y_1, H) - R(Y_2, H)\| \geq 0,
\end{aligned}
\tag{11}
$$

where the equality holds if and only if $F = G$.

The multivariate spatial rank test based on T_M loses the distribution-free property under the null hypothesis. The test relies on the permutation method to determine critical values or compute p-values. But the test is robust. For example, it does not require the assumption of finite second moment as the Hotelling's T^2 test. Neither it requires the assumption of finite first moment as the test (CT) considered by Baringhaus and Franz (2004).

A simulation is conducted to compare performance of T_M, CT and the Hotelling's T^2 under multivariate normal, t_1 and Pareto distributions on $\mathbb{R}^d$ ($d = 2, 5$). Location and scatter alternatives are considered. For location alternatives in normal and t_1 distributions, the parameters of distributions for generating X samples of size $n = 50$ are $\mu = 0$ and $\Sigma_X = I$, while for Y samples with size

Table 6 *Power performance of T_M, CT and Hotelling tests with significance level $\alpha = 0.05$ for multivariate normal, t_1 and Pareto distributions with location alternatives with sample sizes $n = m = 50$*

Dist	Dim	Method	$\Delta = 0$	$\Delta = 0.25$	$\Delta = 0.50$	$\Delta = 0.75$	$\Delta = 1$
Norm	$d = 2$	T_M	0.0550	0.3000	0.8688	0.9966	1
		CT	0.0556	0.3090	0.8818	0.9972	1
		Hotelling	0.0518	0.3226	0.8900	0.9976	1
	$d = 5$	T_M	0.0484	0.5178	0.9944	1	1
		CT	0.0500	0.5332	0.9958	1	1
		Hotelling	0.0494	0.5248	0.9942	1	1
t_1	$d = 2$	T_M	0.0538	0.1574	0.4898	0.8212	0.9644
		CT	0.0596	0.0820	0.226	0.4504	0.7134
		Hotelling	0.0546	0.0562	0.0934	0.1360	0.2058
	$d = 5$	T_M	0.0478	0.2382	0.7986	0.9888	0.9996
		CT	0.0546	0.0858	0.288	0.6200	0.8568
		Hotelling	0.0472	0.0742	0.1622	0.2990	0.4608
Pareto	$d = 2$	T_M	0.0492	0.3470	0.8682	0.9886	0.9998
		CT	0.0560	0.1146	0.2850	0.5330	0.7298
		Hotelling	0.0484	0.0986	0.1858	0.3076	0.4188
	$d = 5$	T_M	0.0522	0.2892	0.7942	0.9784	0.9996
		CT	0.0492	0.1142	0.2942	0.5184	0.7128
		Hotelling	0.0528	0.1108	0.2614	0.4462	0.6046

$m = 50$ are $\boldsymbol{\mu} = (\Delta, \ldots, \Delta)^T$ and $\boldsymbol{\Sigma}_Y = \boldsymbol{I}$, where $\Delta = 0, 0.25, 0.5, 0.75$ and 1. For Pareto distribution, $\boldsymbol{X} = (X_1, \ldots, X_d)^T$ is generated with each component X_j from Pareto(1, 1) and $\boldsymbol{Y} = (Y_1, \ldots, Y_d)^T$ is generated with each component Y_j from Pareto($1 + \Delta$, 1). R package "Hotelling" is used for the Hotelling's T^2 test. T_M and CT tests use the permutation method to compute p-values and $M = 10,000$ iterations are computed to estimate powers by calculating the fraction of p-values less than or equal 0.05. Results for the location alternatives are listed in Table 6.

From Table 6, three tests keep the size 5% well. Powers in $d = 5$ are higher than that in $d = 2$ for each of three tests under all distributions. In the normal cases, T_M performs slightly worse than the Hotelling's test and CT. The power of T_M is about 2% lower than that of the Hotelling test and 1% lower than that of CT under H_a when $\Delta = 0.25$ and $\Delta = 0.50$. However, the power gain of T_M over CT and the Hotelling's test is huge in the t_1-distributions. For $\Delta = 0.25$ and $\Delta = 0.5$, T_M is about twice powerful as CT and about triple powerful as the Hotelling test. The advantage of our proposed T_M over CT and the Hotelling's test are even more significant in the asymmetric Parato distributions than in the t_1 distributions for the location alternatives.

Results for scatter alternatives are listed in Table 7. For multivariate normal and t_1 distributions, we first consider the difference of scatter matrix only on scales.

Table 7 *Power performance of T_M, CT and Hotelling tests with significance level $\alpha = 0.05$ for multivariate normal, t_1 and Pareto distributions with Scatter alternatives with sample sizes $n = m = 50$*

Dist	Dim	Method	$\Delta = 1$	$\Delta = 1.5$	$\Delta = 2$	$\Delta = 2.5$	$\Delta = 3$	Orient
Norm	$d = 2$	T_M	0.0468	0.0640	0.1064	0.1902	0.2992	0.3179
		CT	0.0474	0.0982	0.2716	0.5660	0.8072	0.3016
		Hotelling	0.048	0.0472	0.0598	0.0496	0.0524	0.0493
	$d = 5$	T_M	0.0476	0.0748	0.1272	0.2600	0.4124	0.9678
		CT	0.0470	0.1510	0.5580	0.9192	0.9948	0.8188
		Hotelling	0.045	0.0568	0.0526	0.0576	0.0540	0.0538
t_1	$d = 2$	T_M	0.0486	0.0580	0.0698	0.0948	0.1256	0.2366
		CT	0.0482	0.0680	0.1182	0.1754	0.2370	0.0916
		Hotelling	0.0506	0.0476	0.0488	0.0530	0.0544	0.0495
	$d = 5$	T_M	0.0514	0.0648	0.0900	0.1286	0.1742	0.5896
		CT	0.0512	0.0836	0.1510	0.2320	0.3344	0.1984
		Hotelling	0.0528	0.0494	0.0492	0.0556	0.0550	0.0468
Pareto	$d = 2$	T_M	0.0550	0.6164	0.9802	1	1	–
		CT	0.0540	0.6896	0.9896	1	1	–
		Hotelling	0.0498	0.5148	0.9354	0.9876	0.9976	–
	$d = 5$	T_M	0.0504	0.9158	1	1	1	–
		CT	0.0512	0.9268	0.9996	1	1	–
		Hotelling	0.0566	0.7616	0.9960	0.9998	1	–

The parameters for **X** sample are $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma}_X = \mathbf{I}$, while for **Y** samples are $\boldsymbol{\mu} = \mathbf{0}$ and and $\boldsymbol{\Sigma}_Y = \Delta \mathbf{I}$, where $\Delta = 1, 1.5, 2, 2.5$ and 3. We then consider the alternative with different orientation on the scatter matrices. The scatter matrix is $\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$ for **X** samples, while it is $\begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}$ for **Y** samples. Hence, two components of **X** are positively correlated and the two components of **Y** are negatively correlated. The results for orientation difference alternatives are listed in the last column "Orient" of Table 7. In $d = 5$, $\boldsymbol{\Sigma}_X$ has diagonal elements to be 1 and off-diagonal elements to be 0.5 and $\boldsymbol{\Sigma}_Y$ is constructed to have the same eigenvectors as $\boldsymbol{\Sigma}_X$ and eigenvalues to be the reciprocals of eigenvalues of $\boldsymbol{\Sigma}_X$. For Pareto distributions, $\boldsymbol{X} = (X_1, \ldots, X_d)^T$ is generated with each component X_j from Pareto(1, 1) and $\boldsymbol{Y} = (Y_1, \ldots, Y_d)^T$ is generated with each component Y_j from Pareto(1, Δ).

From Table 7, all tests maintain the size 5% well. For asymmetric Pareto distributions, CT is slightly better than T_M and T_M is better than the Hotelling's test. For normal and t_1 distributions, the Hotelling's T^2 completely fails in scatter alternatives since it is a test on location difference. CT test is much better than T_M for scale alternatives. Particularly CT is triple powerful as the T_M in normal case and twice powerful in the t_1 case. This result is not surprising since T_M is based on the spatial ranks that lose major information on distances or scales. However, when two scatter matrices of distributions are different on orientation, T_M performs bet-

ter than CT, especially in t_1 distribution, the power of T_M is twice or triple as that of CT.

5 Summary

The problem of testing whether two samples come from the same or different population is a classical one in statistics. In this paper, we have studied a rank-based test for the univariate two sample problem. The test statistic is the difference between the average of between-group rank distances and the average of within-group rank distances. Under the null hypothesis, it is distribution free. The limiting null distribution was explored through techniques of Hájek projection and orthogonal decomposition. It has been proved that the limiting distribution is not normal since the projection on one variable is insufficient to represent the variation of the test statistic. By taking the second-order projection, an operator in the functional space was defined and its eigenfunctions and eigenvalues were applied to derive the limiting distribution. It is a weighted mixture of independent chi-square distributions with the weights being the eigenvalues of the operator. We provided a recommendation how to use the limiting distribution to obtain critical values of the proposed test in practice.

The proposed rank test statistic is closely related to two sample Cramér–von Mises criterion. Both statistics are different sample plug-in forms from the same population quantity. We have provided a counter example to show they are different. However, they have the same expectation, variance and limiting distribution. The advantage of our rank test is to allow straightforward generalizations to the multivariate case by using different multivariate rank functions. A continuation of this work is to study properties of the multivariate Cramér–von Mises T_M test. Also the generalizations based on other multivariate rank functions deserve further investigation.

6 Proofs

The following lemma gives the expected value of the absolute difference between the standardized ranks of X and Y.

Lemma 6.1. *Let X and Y be independent continuous random variables from F and G, respectively. Let $H = \tau F + (1 - \tau)G$ with $0 \le \tau \le 1$ be the mixture distribution, J be the distribution of $R(X, H)$ and K be the distribution function of $R(Y, H)$. Then*

$$\mathbb{E}\big|R(X, H) - R(Y, H)\big| = \int_0^1 J(t)\big(1 - K(t)\big)\,dt + \int_0^1 K(t)\big(1 - J(t)\big)\,dt. \quad (12)$$

Proof of Lemma 6.1. Notice that

$$\big|R(X, H) - R(Y, H)\big|$$
$$= \int_0^1 \big[I\big(R(X, H) \le s < R(Y, H)\big) + I\big(R(Y, H) \le s < R(X, H)\big)\big]\,ds.$$

Since H is continuous and $R(X, H) = H(X)$, $R(Y, H) = H(Y)$, we have $J(x) = F \circ H^{-1}(x)$, $K(x) = G \circ H^{-1}(x)$ for any $x \in [0, 1]$, where $H^{-1}(x) = \inf\{u : H(u) \ge x\}$. Then (12) holds by Fubini's theorem. $\qquad\square$

Proof of Theorem 2.2. Define

$$\tilde{D} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n \big|R(X_i, H) - R(Y_j, H)\big|$$

$$- \frac{1}{2m^2} \sum_{i=1}^m \sum_{j=1}^m \big|R(X_i, H) - R(X_j, H)\big|$$

$$- \frac{1}{2n^2} \sum_{i=1}^n \sum_{j=1}^n \big|R(Y_i, H) - R(Y_j, H)\big|.$$

Conditioning on Y_j, $1 \le j \le n$, by the law of large numbers,

$$\frac{1}{m} \sum_{i=1}^m \big|R(X_i, H) - R(Y_j, H)\big| - \mathbb{E}_{X_1}\big|R(X_1, H) - R(Y_j, H)\big| \to 0 \qquad \text{a.s.}$$

and

$$\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n \big|R(X_i, H) - R(Y_j, H)\big|$$

$$= \frac{1}{n} \sum_{j=1}^n \big[\mathbb{E}_{X_1}\big|R(X_1, H) - R(Y_j, H)\big| + o_{\text{a.s.}}(1)\big]$$

$$= \mathbb{E}\big|R(X, H) - R(Y, H)\big| + o_{\text{a.s.}}(1).$$

By the strong law of large numbers for U-statistics (Hoeffding (1961)),

$$\frac{1}{2m^2} \sum_{i=1}^m \sum_{j=1}^m \big|R(X_i, H) - R(X_j, H)\big| \to \frac{1}{2} \mathbb{E}\big|R(X_1, H) - R(X_2, H)\big| \qquad \text{a.s.}$$

and

$$\frac{1}{2n^2} \sum_{i=1}^n \sum_{j=1}^n \big|R(Y_i, H) - R(Y_j, H)\big| \to \frac{1}{2} \mathbb{E}\big|R(Y_1, H) - R(Y_2, H)\big| \qquad \text{a.s.}$$

Then $\tilde{D} \to D$ a.s. as $m, n \to \infty$. Now we show $\hat{D} - \tilde{D} \to 0$ a.s.

By Glivenko–Cantelli theorem, for $m/N \to \tau$,

$$R(x, H_N) \to R(x, H) \qquad \text{a.s.}$$

uniformly on $x \in \mathbb{R}$. Then

$$\left| R(x, H_N) - R(y, H_N) \right| \to \left| R(x, H) - R(y, H) \right| \qquad \text{a.s.}$$

uniformly on $x, y \in \mathbb{R}$ and therefore,

$$\hat{D} - \tilde{D} = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \{ \left| R(X_i, H_N) - R(Y_j, H_N) \right| - \left| R(X_i, H) - R(Y_j, H) \right| \}$$

$$- \frac{1}{2m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} \{ \left| R(X_i, H_N) - R(X_j, H_N) \right| - \left| R(X_i, H) - R(X_j, H) \right| \}$$

$$- \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \{ \left| R(Y_i, H_N) - R(Y_j, H_N) \right| - \left| R(Y_i, H) - R(Y_j, H) \right| \}.$$

$$\to 0 \qquad \text{a.s.}$$

This completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Since all sample ranks are computed with respect to the combined sample, for simple presentation, we write the standardized rank $R(X_i, H_N)$ as $R(X_i)$. We also denote the natural rank of X_i as R_i, that is, $R(X_i) = R_i/N$.

Proof of Theorem 2.4. Under H_0, $R(X_i)$ and $R(Y_j)$ are identically distributed from the discrete uniform distribution on $\{1/N, 2/N, \ldots, (N-1)/N, 1\}$ for all i's and j's. Hence,

$$\mathbb{E}T = \frac{mn}{N} \left\{ \mathbb{E} \left| R(X_1) - R(Y_1) \right| - \frac{m-1}{2m} \mathbb{E} \left| R(X_1) - R(X_2) \right| \right.$$

$$\left. - \frac{n-1}{2n} \mathbb{E} \left| R(Y_1) - R(Y_2) \right| \right\}$$

$$= \frac{1}{2} \mathbb{E} \left| R(X_1) - R(X_2) \right| = \frac{1}{2N} \mathbb{E} |R_1 - R_2|,$$

where R_1 and R_2 are natural ranks of X_1 and X_2, respectively. Under H_0, R_1 and R_2 are two samples drawn uniformly from $\{1, 2, \ldots, N\}$ without replacement. Clearly,

$$\mathbb{E}|R_1 - R_2| = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} |i - j| = \frac{N+1}{3}. \qquad (13)$$

Hence $\mathbb{E}T = \frac{N+1}{6N}$.

Let R_1, R_2, R_3, R_4 be natural ranks of X_1, X_2, X_3 and X_4, respectively. We have

$$\mathbb{E}(R_1 - R_2)^2 = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} (i-j)^2 = \frac{N(N+1)}{6}, \tag{14}$$

$\mathbb{E}|R_1 - R_2||R_1 - R_3|$

$$= \frac{1}{N(N-1)(N-2)} \sum_{i=1}^{N} \sum_{j \neq i} \sum_{k \neq i,j} |i-j||i-k|$$

$$= \frac{1}{N(N-1)(N-2)} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k \neq j} |i-j||i-k| \tag{15}$$

$$= \frac{(N+1)(7N+4)}{60},$$

$\mathbb{E}|R_1 - R_2||R_3 - R_4|$

$$= \frac{1}{N(N-1)(N-2)(N-3)} \sum_{i=1}^{N} \sum_{j \neq i, j=1}^{N} \sum_{k \neq i,j,k=1}^{N} \sum_{l \neq i,j,k;l=1}^{N} |i-j||k-l|$$

$$= \frac{1}{N(N-1)(N-2)(N-3)} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k \neq i,j,k=1}^{N} \sum_{l \neq i,j,k;l=1}^{N} |i-j||k-l| \tag{16}$$

$$= \frac{(N+1)(5N+4)}{45}.$$

Now extending T^2 and $\mathbb{E}T^2$ yields the above three types of expectations $\mathbb{E}(R_1 - R_2)^2$, $\mathbb{E}|R_1 - R_2||R_1 - R_3|$ and $\mathbb{E}|R_1 - R_2||R_3 - R_4|$ denoted as E_1, E_2, E_3, respectively. That is,

$$\mathbb{E}T^2 = \frac{m^2 n^2}{N^4} \left\{ \left[\frac{mn}{m^2 n^2} + \frac{2m(m-1)}{4m^2} + \frac{2n(n-1)}{4n^4} \right] E_1 \right.$$

$$+ \left[\frac{mn(m-1)(n-1)}{m^2 n^2} \right.$$

$$+ \frac{m(m-1)(m-2)(m-3)}{4m^4} + \frac{n(n-1)(n-2)(n-3)}{4n^4}$$

$$- \frac{2mn(m-1)(m-2)}{2m^3 n} - \frac{2mn(n-1)(n-2)}{2mn^3} + \left. \frac{2m(m-1)n(n-1)}{4m^2 n^2} \right] E_3$$

$$+ \left[\frac{m^2 n^2 - mn - mn(m-1)(n-1)}{m^2 n^2} + \frac{4m(m-1)(m-2)}{4m^4} \right.$$

$$+ \frac{4n(n-1)(n-2)}{4n^4}$$

$$- \frac{m^2 n(m-1) - mn(m-1)(m-2)}{m^3 n}$$

$$- \left. \frac{mn^2(n-1) - mn(n-1)(n-2)}{mn^3} \right] E_2 \Big\}$$

$$= \frac{N+1}{60N^2} \left[3N + 3 - \frac{N^2}{mn} \right].$$

Hence,

$$\mathrm{Var}(T) = \mathbb{E}T^2 - (\mathbb{E}T)^2 = \frac{N+1}{180N^2} \left[4(N+1) - \frac{3N^2}{mn} \right].$$

This completes the proof.   □

Lemma 6.2. *Under H_0, $\mathbb{E}[|R(X_1) - R(X_2)||X_1] = \frac{1}{2} - \frac{N-2}{N}[F(X_1) - F^2(X_1)]$, $\mathbb{E}[|R(X_2) - R(X_3)||X_1] = \frac{1}{3} + \frac{2}{N}[F(X_1) - F^2(X_1)]$ and hence $\mathbb{E}(T|X_1) = \frac{1}{6}(1 + \frac{n}{mN}) + \frac{1}{N}(1 - \frac{n}{m})[F(X_1) - F^2(X_1)]$.*

Proof of Lemma 6.2. Let R_1 be the natural rank of X_1 and R_2 be the natural rank of X_2. Under H_0 and given X_1, $R_1 - 1$ has a binomial distribution with parameters $N - 1$ and $F(X_1)$, that is,

$$P(R_1 - 1 = u|X_1) = \binom{N-1}{u} F(X_1)^u [1 - F(X_1)]^{N-1-u},$$

$$u = 0, 1, \ldots, N - 1.$$

Given R_1, R_2 is uniformly distributed from $\{1, 2, \ldots, N\}/\{R_1\}$. So

$$\mathbb{E}(|R_1 - R_2||X_1)$$

$$= \mathbb{E}[\mathbb{E}(|R_1 - R_2||R_1)|X_1] = \sum_{r_1=1}^{N} \frac{1}{N-1} \left[\sum_{i=1, \neq r_1}^{N} |r_1 - i| \right] P(R_1 = r_1|X_1)$$

$$= \frac{1}{2(N-1)} \sum_{r_1=1}^{N} [(N-1)N - 2(N-1)(r_1 - 1) + 2(r_1 - 1)^2]$$

$$\times P(R_1 = r_1|X_1)$$

$$= \frac{1}{2}N - (N-1)F(X_1)$$

$$+ \frac{1}{N-1}[(N-1)F(X_1)(1 - F(X_1) + (N-1)^2 F^2(X_1)]$$

$$= \frac{1}{2}N - (N-2)[F(X_1) - F^2(X_1)].$$

Let R_3 be the natural rank of X_3. Under H_0, we have

$$\mathbb{E}\big(|R_2 - R_3||X_1\big)$$

$$= \mathbb{E}\big[\mathbb{E}\big(|R_2 - R_3||R_1\big)|X_1\big]$$

$$= \sum_{r_1=1}^{N} \frac{1}{(N-1)(N-2)} \left[\sum_{i=1,i\neq r_1}^{N} \sum_{j=1,j\neq i,r_1}^{N} |i-j|\right] P(R_1 = r_1|X_1)$$

$$= \frac{N(N+1)}{3(N-2)} - \frac{N-2(N-2)[F(X_1) - F^2(X_1)]}{N-2}$$

$$= \frac{1}{3}N + 2[F(X_1) - F^2(X_1)].$$

It is clear that $\mathbb{E}(T|X_1)$ contains the above conditional expectations $\mathbb{E}[|R(X_1) - R(X_2)||X_1]$ and $\mathbb{E}[|R(X_2) - R(X_3)||X_1]$ denoted as E_1^* and E_2^*, respectively. Then it follows that

$$\mathbb{E}(T|X_1)$$

$$= \frac{mn}{N}\left\{\frac{n}{mn}E_1^* + \frac{mn-n}{mn}E_2^* - \frac{2(m-1)}{2m^2}E_1^* - \frac{(m-1)(m-2)}{2m^2}E_2^* \right.$$

$$\left. - \frac{n(n-1)}{2n^2}E_2^*\right\}$$

$$= \frac{mn}{N}\left\{\frac{1}{m^2}(E_1^* - E_2^*) + \frac{N}{2mn}E_2^*\right\}$$

$$= \frac{1}{6}\left(1 + \frac{n}{mN}\right) + \frac{1}{N}\left(1 - \frac{n}{m}\right)[F(X_1) - F^2(X_1)].$$

This complete the proof. $\square$

The following two lemmas are on the second order projection of T.

Lemma 6.3. *Under H_0, the second order projection of T on one X variable and one Y variable is*

$$\mathbb{E}[T|X_1, Y_1]$$

$$= \frac{mnN + N^2 - 7mn}{6mnN}$$

$$+ \frac{(2m-n)(n-1)}{mnN}F(X_1)\big[1 - F(X_1)\big]$$

$$+ \frac{(2n-m)(m-1)}{mnN}F(Y_1)\big[1 - F(Y_1)\big]$$

(17)

$$+ I(Y_1 > X_1)\left\{\frac{5n - m}{6mnN} + \frac{m - 1}{mN}F(Y_1) - \frac{n - 1}{nN}F(X_1)\right\}$$

$$+ I(Y_1 < X_1)\left\{\frac{5m - n}{6mnN} - \frac{m - 1}{mN}F(Y_1) + \frac{n - 1}{nN}F(X_1)\right\}$$

and its variance is

$$\text{Var}\{\mathbb{E}[T|X_1, Y_1]\}$$

$$= \frac{m^4 + n^4 - 2m^3n - 2mn^3 + 10m^2n^2 - 8mnN + 5n^2 + 5m^2}{180N^2m^2n^2}. \tag{18}$$

Proof of Lemma 6.3. In the proof of next two lemmas, we use $\mathbb{E}[S|Z_1, Z_2, Z_1 < Z_2]$ to denote $\mathbb{E}[SI(Z_1 < Z_2)|Z_1, Z_2]$ for any random variables S, Z_1 and Z_2. Again, let R_i be the natural rank of X_i, $i = 1, 2, 3, 4$. Under H_0 and given $X_1 < X_2$, (R_1, R_2) has trinomial distribution with parameters $F(X_1)$, $F(X_2) - F(X_1)$ and $1 - F(X_2)$, that is,

$$P(R_1 = u, R_2 = v|X_1, X_2, X_1 < X_2) \tag{19}$$

$$= \binom{N - 2}{u - 1, v - u - 1, N - v}[F(X_1)]^{u - 1} \tag{20}$$

$$\times [F(X_2) - F(X_1)]^{v - u - 1}[1 - F(X_2)]^{N - v}I(X_1 < X_2).$$

Therefore,

$$\mathbb{E}[|R_1 - R_2||X_1, X_2, X_1 < X_2]$$
$$= \{(N - 2)[F(X_2) - F(X_1)] + 1\}I(X_1 < X_2). \tag{21}$$

Under H_0 and given R_1, R_2, the natural rank R_3 of X_3 has discrete uniform distribution on the set $\{1, 2, \ldots, N\}/\{R_1, R_2\}$, that is,

$$P(R_3 = w|R_1, R_2) = \frac{1}{N - 2}$$

for $1 \leq w \leq N$, $w \neq R_1, R_2$. Therefore,

$$\mathbb{E}(|R_1 - R_3||X_1, X_2, X_1 < X_2)$$

$$= \mathbb{E}[\mathbb{E}(|R_1 - R_3||R_1 < R_2)|X_1, X_2, X_1 < X_2]$$

$$= \mathbb{E}\left[\frac{1}{N - 2}\left(\sum_{1 \leq i < R_1}(R_1 - i) + \sum_{R_1 < i \leq N, i \neq R_2}(i - R_1)\right)\right|$$

$$X_1, X_2, X_1 < X_2\right] \tag{22}$$

$$= \mathbb{E}\left(\frac{N^2 + N - 2NR_1 + 2R_1^2 - 2R_2}{2(N - 2)}\Big|X_1, X_2, X_1 < X_2\right)$$

$$= \left\{ \frac{N+1}{2} - (N-3)F(X_1)[1 - F(X_1)] - F(X_2) \right\} I(X_1 < X_2).$$

The last equality (22) is from (19), the conditional distribution of (R_1, R_2). By a similar calculation,

$$\mathbb{E}(|R_1 - R_3||X_1, X_2, X_1 > X_2)$$

$$= \left\{ \frac{N-1}{2} - (N-3)F(X_1)[1 - F(X_1)] + F(X_2) \right\} I(X_1 > X_2). \tag{23}$$

We also have

$$\mathbb{E}(|R_3 - R_4||X_1, X_2, X_1 < X_2)$$

$$= \mathbb{E}[\mathbb{E}(|R_3 - R_4||R_1 < R_2)|X_1, X_2, X_1 < X_2]$$

$$= \mathbb{E}\left[\frac{1}{(N-2)(N-3)} \sum_{1 \le i, j \le N, i, j \ne R_1, R_2} |i - j| \Big| X_1, X_2, X_1 < X_2 \right] \tag{24}$$

$$= \left\{ \frac{N-1}{3} + 2F(X_1)[1 - F(X_1)] + 2F(X_2)[1 - F(X_2)] \right\} I(X_1 < X_2).$$

Again, the last equality (24) is from (19), the conditional distribution of (R_1, R_2). Now let R_1, R_2 be the natural rank of X_1 and Y_1, respectively. R_3 and R_4 be the natural ranks of two other different X_i or Y_j, $1 < i \le m, 1 < j \le n$. By the definition of T as in (4),

$$\frac{N}{mn} \mathbb{E}[T|X_1, Y_1, X_1 < Y_1]$$

$$= \frac{1}{mn} \{ \mathbb{E}[|R(X_1) - R(Y_1)||X_1, Y_1, X_1 < Y_1]$$

$$+ (n-1)\mathbb{E}(|R(X_1) - R(Y_2)||X_1, Y_1, X_1 < Y_1)$$

$$+ (m-1)\mathbb{E}(|R(X_2) - R(Y_1)||X_1, Y_1, X_1 < Y_1)$$

$$+ (m-1)(n-1)\mathbb{E}(|R(X_2) - R(Y_2)||X_1, Y_1, X_1 < Y_1) \}$$

$$- \frac{1}{2m^2} \{ 2(m-1)\mathbb{E}(|R(X_1) - R(X_2)||X_1, Y_1, X_1 < Y_1)$$

$$+ (m-1)(m-2)\mathbb{E}(|R(X_2) - R(X_3)||X_1, Y_1, X_1 < Y_1) \}$$

$$- \frac{1}{2n^2} \{ 2(n-1)\mathbb{E}(|R(Y_1) - R(Y_2)||X_1, Y_1, X_1 < Y_1)$$

$$+ (n-1)(n-2)\mathbb{E}(|R(Y_2) - R(Y_3)||X_1, Y_1, X_1 < Y_1) \}$$

$$= \frac{1}{mnN} \mathbb{E}(|R_1 - R_2||X_1, Y_1, X_1 < Y_1)$$

$$+ \frac{n-m}{m^2nN}\mathbb{E}\big(|R_1 - R_3| \,\big|\, X_1, Y_1, X_1 < Y_1\big)$$

$$+ \frac{m-n}{mn^2N}\mathbb{E}\big(|R_2 - R_3| \,\big|\, X_1, Y_1, X_1 < Y_1\big)$$

$$+ \frac{mnN + 6mn - 2N^2}{2m^2n^2N}\mathbb{E}\big(|R_3 - R_4| \,\big|\, X_1, Y_1, X_1 < Y_1\big).$$

By (21), (22), (23) and (24), we have

$$\mathbb{E}[T|X_1, Y_1, X_1 < Y_1]$$

$$= I(Y_1 > X_1)\bigg\{ \frac{1}{N^2}\big\{(N-2)[F(Y_1) - F(X_1)] + 1\big\}$$

$$+ \frac{n-m}{mN^2}\bigg\{ \frac{N+1}{2} - (N-3)F(X_1)[1 - F(X_1)] - F(Y_1)\bigg\}$$

$$+ \frac{m-n}{nN^2}\bigg\{ \frac{N-1}{2} - (N-3)F(Y_1)[1 - F(Y_1)] + F(X_1)\bigg\}$$

$$+ \frac{mnN + 6mn - 2N^2}{2mnN^2}$$

$$\times \bigg\{ \frac{N-1}{3} + 2F(X_1)[1 - F(X_1)] + 2F(Y_1)[1 - F(Y_1)]\bigg\}\bigg\}.$$

Hence,

$$\mathbb{E}[T|X_1, Y_1, X_1 < Y_1]$$

$$= I(Y_1 > X_1)\bigg\{ \frac{mnN + N^2 - 7mn + 5n - m}{6mnN}$$

$$+ \frac{(2m-n)(n-1)}{mnN}F(X_1)[1 - F(X_1)]$$

$$+ \frac{(2n-m)(m-1)}{mnN}F(Y_1)[1 - F(Y_1)]$$

$$+ \frac{m-1}{mN}F(Y_1) - \frac{n-1}{nN}F(X_1)\bigg\}.$$

A similar calculation gives

$$\mathbb{E}[T|X_1, Y_1, X_1 > Y_1]$$

$$= I(Y_1 < X_1)\bigg\{ \frac{mnN + N^2 - 7mn + 5m - n}{6mnN}$$

$$+ \frac{(2m-n)(n-1)}{mnN}F(X_1)[1 - F(X_1)]$$

$$+ \frac{(2n-m)(m-1)}{mnN} F(Y_1)[1 - F(Y_1)]$$

$$- \frac{m-1}{mN} F(Y_1) + \frac{n-1}{nN} F(X_1) \Big\}.$$

Therefore, (17) holds.

Let U_1, U_2 be i.i.d. uniform random variables on [0, 1], then

$$\mathrm{Var}\{\mathbb{E}[T | X_1, Y_1]$$

$$= \mathrm{Var}\Big\{ \frac{(2m-n)(n-1)}{mnN} U_1(1 - U_1) + \frac{(2n-m)(m-1)}{mnN} U_2(1 - U_2)$$

$$+ I(U_2 > U_1)\Big\{ \frac{5n-m}{6mnN} + \frac{m-1}{mN} U_2 - \frac{n-1}{nN} U_1 \Big\}$$

$$+ I(U_2 < U_1)\Big\{ \frac{5m-n}{6mnN} - \frac{m-1}{mN} U_2 + \frac{n-1}{nN} U_1 \Big\}\Big\}.$$

Therefore, (18) holds. This completes the proof. $\qquad\square$

The following lemma gives the second order projection of T on two X variables or two Y variables.

Lemma 6.4. *Under H_0, the projection of T on two X variables is*

$$\mathbb{E}[T | X_1, X_2]$$

$$= \frac{n}{N}\Big\{ \frac{mN + 6n - m}{6mn}$$

$$+ \frac{m - 2n}{mn}\big[F(X_1)(1 - F(X_1)) + F(X_2)(1 - F(X_2))\big] \tag{25}$$

$$- \frac{1}{m}|F(X_2) - F(X_1)|\Big\}.$$

The variance of the projection is

$$\mathrm{Var}\{\mathbb{E}[T | X_1, X_2]\} = \frac{m^2 - 2mn + 5n^2}{90m^2N^2}. \tag{26}$$

The projection of T on two Y variables is

$$\mathbb{E}[T | Y_1, Y_2]$$

$$= \frac{m}{N}\Big\{ \frac{nN + 6m - n}{6mn} + \frac{n - 2m}{mn}\big[F(Y_1)(1 - F(Y_1)) + F(Y_2)(1 - F(Y_2))\big]$$

$$- \frac{1}{n}|F(Y_2) - F(Y_1)|\Big\}.$$

The variance of the projections is

$$\mathrm{Var}\{\mathbb{E}[T|Y_1, Y_2]\} = \frac{n^2 - 2mn + 5m^2}{90n^2 N^2}. \tag{27}$$

Proof of Lemma 6.4. By symmetry, we only need prove the results for the projection on two X variables. By the definition of T as in (4), under the null hypothesis,

$$\mathbb{E}[T|X_1, X_2, X_1 < X_2]$$

$$= \frac{mn}{N}\left\{ \frac{2}{m^2 N}\mathbb{E}[|R_1 - R_3||X_1, X_2, X_1 < X_2] \right.$$

$$+ \frac{2}{m^2 N}\mathbb{E}[|R_2 - R_3||X_1, X_2, X_1 < X_2]$$

$$+ \frac{mN - 6n}{2m^2 nN}\mathbb{E}[|R_3 - R_4||X_1, X_2, X_1 < X_2]$$

$$\left. - \frac{1}{m^2 N}\mathbb{E}[|R_1 - R_2||X_1, X_2, X_1 < X_2] \right\}.$$

By (21), (22), (23) and (24), we have

$$\mathbb{E}[T|X_1, X_2, X_1 < X_2]$$

$$= \frac{n}{N}\left\{ \frac{mN + 6n - m}{6mn} + \frac{m - 2n}{mn}[F(X_1)(1 - F(X_1)) + F(X_2)(1 - F(X_2))] \right.$$

$$\left. - \frac{1}{m}[F(X_2) - F(X_1)] \right\}I(X_1 < X_2).$$

By symmetry, we have (25). Let U_1, U_2 be i.i.d. uniform random variables on $[0, 1]$, then

$$\mathrm{Var}\{\mathbb{E}[T|X_1, X_2]\}$$

$$= \frac{n^2}{N^2}\mathrm{Var}\left\{ \frac{m - 2n}{mn}[U_1(1 - U_1) + U_2(1 - U_2)] - \frac{1}{m}|U_2 - U_1| \right\}$$

$$= \frac{m^2 - 2mn + 5n^2}{90m^2 N^2}.$$

This completes the proof.   □

In the following, we provide a lemma which is useful in deriving the asymptotics of the test statistic T.

Lemma 6.5. *Let $S_n(X_1, X_2, \ldots, X_n)$ be a function of n independent random variables with decomposition $S_n = M_n + R_n$. If $\mathbb{E}(R_n) = \mathrm{Cov}(M_n, R_n) = 0$ for any n and $\mathrm{Var}(S_n)/\mathrm{Var}(M_n) \to 1$ as $n \to \infty$, then $|R_n|/\sqrt{\mathrm{Var}(S_n)} \to 0$ in $\mathcal{L}^2$ norm and therefore $|R_n|/\sqrt{\mathrm{Var}(S_n)} \to 0$ in probability.*

Proof of Lemma 6.5.

$$\mathbb{E}\big[R_n^2/\operatorname{Var}(S_n)\big]$$

$$= \frac{\mathbb{E}[(S_n - \mathbb{E}S_n) - (M_n - \mathbb{E}M_n)]^2}{\operatorname{Var}(S_n)}$$

$$= \frac{\operatorname{Var}(S_n) + \operatorname{Var}(M_n) - 2\mathbb{E}(S_n - \mathbb{E}S_n)(M_n - \mathbb{E}M_n)}{\operatorname{Var}(S_n)}$$

$$= \frac{\operatorname{Var}(S_n) + \operatorname{Var}(M_n) - 2\mathbb{E}(M_n - \mathbb{E}M_n)^2 - 2\mathbb{E}R_n(M_n - \mathbb{E}M_n)}{\operatorname{Var}(S_n)}$$

$$= \frac{\operatorname{Var}(S_n) - \operatorname{Var}(M_n)}{\operatorname{Var}(S_n)} = 1 - \operatorname{Var}(M_n)/\operatorname{Var}(S_n) \to 0. \qquad \square$$

The above Lemma 6.5 is a result of Hájeck projection technique. See Hájek and Šidák (1967) and Hettmansperger and McKean (2010) for details.

Proof of Theorem 2.5. We can write $h(x, y) = \sum_{k=1}^{\infty} \lambda_k \phi_k(x)\phi_k(y)$, where $\{\phi_k(\cdot)\}$ are the orthonormal eigenfunctions corresponding to the eigenvalues $\{\lambda_k\}$, see Dunford and Schwartz (1963), Serfling (1980). Since $\mathbb{E}h(X, y) = 0$ and

$$0 = \operatorname{Var}\{\mathbb{E}[h(X, Y)|Y]\} = \sum_{k=1}^{\infty} \lambda_k^2 (\mathbb{E}\phi_k(X))^2 \operatorname{Var}(\phi_k(Y)).$$

Therefore, $\mathbb{E}(\phi_k(X)) = 0$ for all $k \geq 1$ and $\operatorname{Var}[h(X, Y)] = \mathbb{E}h^2(X, Y) = \sum_{k=1}^{\infty} \lambda_k^2$. By (8),

$$\sum_{k=1}^{\infty} \lambda_k^2 = \operatorname{Var}[h(X, Y)] = 2/45.$$

The above results can also be confirmed by $\lambda_k = -\frac{2}{\pi^2 k^2}$, $\phi_k(x) = \cos k\pi F(x)$, $k \in \mathbb{N}$. Denote $W_{mk}(X) = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} \phi_k(X_i)$, $W_{nk}(Y) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \phi_k(Y_j)$, $Z_{mk}(X) = \frac{1}{m} \sum_{i=1}^{m} \phi_k^2(X_i)$, $Z_{nk}(Y) = \frac{1}{n} \sum_{j=1}^{n} \phi_k^2(Y_j)$. Define T_{NK} by

$$\sqrt{\operatorname{Var}(T)} T_{NK} = \frac{1}{N} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{K} \lambda_k \phi_k(X_i)\phi_k(Y_j)$$

$$- \frac{1}{N} \sum_{1 \leq i < j \leq m} \sum_{k=1}^{K} \lambda_k \phi_k(X_i)\phi_k(X_j)$$

$$- \frac{1}{N} \sum_{1 \leq i < j \leq n} \sum_{k=1}^{K} \lambda_k \phi_k(Y_i)\phi_k(Y_j).$$

Then

$$\sqrt{\mathrm{Var}(T)}\,T_{NK}$$

$$= \frac{\sqrt{mn}}{N} \sum_{k=1}^{K} \lambda_k W_{mk}(X) W_{nk}(Y) \tag{28}$$

$$- \frac{m}{N} \sum_{k=1}^{K} \frac{\lambda_k}{2} \{W_{mk}^2(X) - Z_{mk}(X)\} - \frac{n}{N} \sum_{k=1}^{K} \frac{\lambda_k}{2} \{W_{nk}^2(Y) - Z_{nk}(Y)\}.$$

Applying the argument as in Serfling (1980), page 197, $|\mathbb{E}e^{ixT_N} - \mathbb{E}e^{ixT_{NK}}| \leq |x| [\mathbb{E}(T_N - T_{NK})^2]^{1/2}]$. It is easy to see that

$$\mathbb{E}(T_N - T_{NK})^2 = \left(\frac{mn}{N^2} + \frac{m(m+1) + n(n-1)}{2N^2} \right) \frac{1}{\mathrm{Var}(T)} \sum_{k=K+1}^{\infty} \lambda_k^2$$

$$= \left[\frac{45}{2} + o(1) \right] \sum_{k=K+1}^{\infty} \lambda_k^2 \leq 23 \sum_{k=K+1}^{\infty} \lambda_k^2.$$

For a given $\varepsilon > 0$ and fixed x, we can choose and fix K to be large enough so that

$$|x| \left(23 \sum_{k=K+1}^{\infty} \lambda_k^2 \right)^{1/2} < \varepsilon.$$

Then we have $|\mathbb{E}e^{ixT_N} - \mathbb{E}e^{ixT_{NK}}| < \varepsilon$ for all N. By (28), Theorem 2.4 and the condition $\lim_{N \to \infty} \frac{m}{n} = 1$, we may write

$$T_{NK} = \frac{\sqrt{45}}{2} \sum_{k=1}^{K} \lambda_k \{-[W_{nk}(X) - W_{nk}(Y)]^2/2 + Z_{nk}(X)/2 + Z_{nk}(Y)/2\} + r_N.$$

It is easy to see that $r_N \to 0$ in probability. Let χ_{1k}^2 be i.i.d. χ_1^2 random variables. Denote $U_K = \frac{\sqrt{45}}{2} \sum_{k=1}^{K} \lambda_k(-\chi_{1k}^2 + 1)$ and $U = \frac{\sqrt{45}}{2} \sum_{k=1}^{\infty} \lambda_k(-\chi_{1k}^2 + 1)$. Since $\mathbb{E}W_{mk}(X) = \mathbb{E}W_{nk}(Y) = 0$ and $W_{mk}(X), W_{nk}(Y)$ are orthonormal, then the random vector $\{(W_{mk}(X) - W_{nk}(Y))/\sqrt{2}\}_{k=1}^{K} \Rightarrow N(0, I_{K \times K})$ by the Linderberg–Lévy central limit theorem, $Z_{mk}(X) \to 1$ and $Z_{nk}(Y) \to 1$ for $1 \leq k \leq K$ by the strong law of large numbers. Hence, $T_{NK} \Rightarrow U_K$ as $n \to \infty$ and $\mathbb{E}(e^{ixT_{NK}}) - \mathbb{E}(e^{ixU_K})| < \varepsilon$ for all N. By the same argument as in Serfling (1980), $|\mathbb{E}(e^{ixU_K}) - \mathbb{E}(e^{ixU})| < \varepsilon$ for all N. Then together with $|\mathbb{E}e^{ixT_N} - \mathbb{E}e^{ixT_{NK}}| < \varepsilon$, we have $|\mathbb{E}e^{ixT_N} - \mathbb{E}e^{ixU}| \to 0$ as $N \to \infty$. Therefore, $T_N \Rightarrow -\frac{\sqrt{45}}{2} \sum_{k=1}^{\infty} \lambda_k(\chi_{1k}^2 - 1)$, where $\chi_{11}^2, \chi_{12}^2, \ldots$ are independent χ_1^2 variables. Since $\mathrm{Var}(\hat{T})/\mathrm{Var}(T) \to 1$, we have $(T - \mathbb{E}T - \hat{T})/\sqrt{\mathrm{Var}(T)} \to 0$ in probability. Then by Lemma 6.5,

$$(T - \mathbb{E}T)/\sqrt{\mathrm{Var}(T)} \Rightarrow -\frac{\sqrt{45}}{2} \sum_{k=1}^{\infty} \lambda_k(\chi_{1k}^2 - 1).$$

This completes the proof.   □

Proof of Theorem 4.1. Let μ be the uniform distribution on the surface of the unit ball $S^{d-1} = \{\boldsymbol{a} \in \mathbb{R}^d : \|\boldsymbol{a}\| = 1\}$. From Theorem 2.1, we have

$$\mathbb{E}\big|R(\boldsymbol{a}^T X, H^a) - R(\boldsymbol{a}^T Y, H^a)\big| - \frac{1}{2}\mathbb{E}\big|R(\boldsymbol{a}^T X_1, H^a) - R(\boldsymbol{a}^T X_2, H^a)\big|$$

$$- \frac{1}{2}\mathbb{E}\big|R(\boldsymbol{a}^T Y_1, H^a) - R(\boldsymbol{a}^T Y_2, H^a)\big| \geq 0$$

for each $\boldsymbol{a} \in S^{d-1}$, where $H^a = \tau F^a + (1 - \tau)G^a$ with F^a and G^a being the distributions of $\boldsymbol{a}^T X$ and $\boldsymbol{a}^T Y$, respectively. Integration of $\boldsymbol{a}$ with respect to μ obtains (11). Equality holds if and only if for μ-almost all $\boldsymbol{a} \in S^{d-1}$ the distributions of $\boldsymbol{a}^T X$ and $\boldsymbol{a}^T Y$ coincide. For each $t \in \mathbb{R}$ the functions $\mathbb{E}\exp(it\boldsymbol{a}^T X)$ and $\mathbb{E}\exp(it\boldsymbol{a}^T Y)$ with $\boldsymbol{a} \in S^{d-1}$ are continuous. Thus, equality in (11) holds if and only if X and Y have the same characteristic function, hence have the same distribution. $\qquad\square$

References

Albers, W., Kallenberg, W. C. M. and Martini, F. (2001). Data-driven rank tests for classes of tail alternatives. *Journal of the American Statistical Association* **96**, 685–696. MR1946434

Anderson, T. W. (1962). On the distribution of the two-sample Cramér–von Mises criterion. *The Annals of Mathematical Statistics* **33**, 1148–1159. MR0145607

Baringhaus, L. and Franz, C. (2004). On a new multivariate two-sample test. *Journal of Multivariate Analysis* **88**, 190–206. MR2021870

Baumgartner, W., Weiß, P. and Schindler, H. (1998). A nonparametric test for the general two sample problem. *Biometrics* **54**, 1129–1135.

Borroni, C. G. (2001). Some notes about the nonparametric tests for the equality of two populations. *Test* **10**, 147–159. MR1856189

Cao, R. and Van Keilegom, I. (2006). Empirical likelihood tests for two-sample problems via non-parametric density estimation. *Canadian Journal of Statistics* **34**, 61–77. MR2267710

Chiu, S. and Liu, K. (2009). Generalized Cramér–von Mises goodness-of-fit tests for multivariate distributions. *Computational Statistics & Data Analysis* **53**, 3817–3834. MR2749926

Cotterill, D. and Csörgő, M. (1982). On the limiting distribution of and critical values for the multi-variate Cramér–von Mises statistic. *The Annals of Statistics* **10**, 233–244. MR0642735

Darling, D. A. (1957). The Kolomogorov–Smirnov, Cramér–von Mises tests. *The Annals of Mathe-matical Statistics* **28**, 823–838. MR0093870

Dunford, N. and Schwartz, J. T. (1963). *Linear Operators Part II: Spectral Theory. Self Adjoint Operators in Hilbert Space*. New York–London: John Wiley & Sons. MR0188745

Efron, B. and Stein, C. (1978). The jackknife estimate of variance. Technical Report No. 40, Division of Biostatistics, Stanford University.

Einmahl, J. and McKeague, I. (2003). Empirical likelihood based hypothesis testing. *Bernoulli* **9**, 267–290. MR1997030

Fernández, V., Jimènez Gamerro, M. and Muñoz Garcìa, J. (2008). A test for the two-sample problem based on empirical characteristic functions. *Computational Statistics & Data Analysis* **52**, 3730–3748. MR2427377

Fisz, M. (1960). On a result by M. Rosenblatt concerning the von Mises–Smirnov test. *The Annals of Mathematical Statistics* **31**, 427–429. MR0121918

Genest, C., Quessy, J. F. and Rémillard, B. (2007). Asymptotic local efficiency of Cramér–von Mises tests for multivariate independence. *The Annals of Statistics* **35**, 166–191. MR2332273

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B. and Smola, A. (2008). A kernel method for the two-sample problem. *Journal of Machine Learning Research* **1**, 1–10.

Gurevich, G. and Vexler, A. (2011). A two-sample empirical likelihood ratio test based on samples entropy. *Statistics and Computing* **21**, 657–670. MR2826699

Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests*. San Diego: Academic Press. MR0229351

Hettmansperger, T. P. and McKean, J. W. (2010). *Robust Nonparametric Statistical Methods*, 2nd ed. London: Chapman & Hall. MR1604954

Hoeffding, W. (1961). The strong law of large numbers for U-statistics. Inst. Statist. Univ. of North Carolina, Mimeo Report, No. 302.

Janic-Wróblewska, A. and Ledwina, T. (2000). Data driven rank test for two-sample problem. *Scandinavian Journal of Statistics* **27**, 281–297. MR1777504

Lehmann, E. L. (1951). Consistency and unbiasedness of certain nonparametric tests. *The Annals of Mathematical Statistics* **22**, 165–179. MR0040632

Möttönen, J., Oja, H. and Tienari, J. (1997). On the efficiency of multivariate spatial sign and rank tests. *The Annals of Statistics* **25**, 542–552. MR1439313

Oja, H. (2010). *Multivariate Nonparametric Methods with R: An Approach Based on Spatial Signs and Ranks*. New York: Springer. MR2598854

Pettitt, A. N. (1976). A two-sample Anderson–Darling rank statistic. *Biometrika* **63**, 161–168. MR0413359

Rosenblatt, M. (1952). Limit theorems associated with variants of the von Mises statistic. *The Annals of Mathematical Statistics* **23**, 617–623. MR0052732

Schmid, F. and Trede, M. (1995). A distribution free test for the two sample problem for general alternatives. *Computational Statistics & Data Analysis* **20**, 409–419.

Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*. New York: Wiley. MR0595165

Székely, G. J. and Rizzo, M. L. (2004). Testing for equal distributions in high dimension. *InterStat*, Nov. 5.

Székely, G. J. and Rizzo, M. L. (2013). Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference* **143**, 1249–1272. MR3055745

Székely, G. J. and Rizzo, M. L. (2017). The energy of data. *Annual Review of Statistics and Its Application* **4**, 447–479.

J. Curry
School of Science & Technology
Georgia Gwinnett College
Lawrenceville, Georgia 30043
USA
E-mail: jcurrysavag@ggc.edu

X. Dang
H. Sang
Department of Mathematics
University of Mississippi
University, Mississippi 38677
USA
E-mail: xdang@olemiss.edu
        sang@olemiss.edu