Two-sample test based on maximum variance discrepancy

N. Makigusa
Graduate School of Nanobioscience, Yokohama City University, Yokohama, Japan

ABSTRACT
In this article, we introduce a novel discrepancy called the maximum variance discrepancy for the purpose of measuring the difference between two distributions in Hilbert spaces that cannot be found via the maximum mean discrepancy. We also propose a two-sample goodness of fit test based on this discrepancy. We obtain the asymptotic null distribution of this two-sample test, which provides an efficient approximation method for the null distribution of the test.

ARTICLE HISTORY
Received 22 June 2022
Accepted 22 May 2023

KEYWORDS
Kernel method; maximum mean discrepancy; reproducing kernel Hilbert space; two-sample test

1. Introduction
For probability distributions $P$ and $Q$, the test for the null hypothesis $H_0 : P = Q$ against an alternative hypothesis $H_1 : P \neq Q$ based on data $X_1, \ldots, X_n \overset{i.i.d.}{\sim} P$ and $Y_1, \ldots, Y_m \overset{i.i.d.}{\sim} Q$ is known as a two-sample test. Such tests have applications in various areas. There is a huge body of literature on two-sample tests in Euclidean space, so we will not attempt a complete bibliography. Two-sample tests for high-dimensional data have been researched in Kim, Nam, and Kim (2013) and Zhang, Peng, and Zhang (2010). Kernel-based approaches used for statistical inference have been discussed in Fukumizu et al. (2009); Gretton et al. (2007, 2009, 2012); and Thi Thien Trang et al. (2021). Specifically, in Gretton et al. (2007), a two-sample test based on Maximum Mean Discrepancy (MMD) is proposed. The MMD of the reproducing kernel Hilbert space $H(k)$ associated with the positive-definite kernel $k$ is also defined in Gretton et al. (2007).

In this article, we propose a novel discrepancy between two distributions defined as

$$T = \| V_X \sim P[k(\cdot, X)] - V_Y \sim Q[k(\cdot, Y)] \|_{H(k) \otimes \mathbb{R}^2},$$

and we call this the Maximum Variance Discrepancy (MVD), where $V_X \sim P[k(\cdot, X)]$ is a covariance operator in $H(k)$. The MVD is composed by replacing the kernel mean embedding in MMD with a covariance operator; hence, it is natural to consider a two-sample test based on the MVD. A related work can be found in Boente, Rodriguez, and Sued (2018), where a test for the equality of covariance operators in Hilbert spaces was proposed.

Our aim in this research is to clarify the properties of the MVD test from two perspectives: an asymptotic investigation as $n, m \to \infty$, and its practical implementation. We first obtain the asymptotic distribution of a consistent estimator $\hat{T}_{n,m}^2$ of $T^2$, under $H_0$. We also derive the asymptotic distribution of $\hat{T}_{n,m}^2$ under the alternative hypothesis $H_1$. Furthermore, we consider a sequence of local alternative distributions $Q_{nm} = (1 - 1/\sqrt{n+m})P + (1/\sqrt{n+m})Q$ for $P \neq Q$ and address the asymptotic distribution of $\hat{T}_{n,m}^2$ under this sequence. For practical
purposes, a method to approximate the distribution of the test by \( \hat{T}_{n,m}^2 \) under \( H_0 \) is developed. The method is based on the eigenvalues of the centered Gram matrices associated with the dataset. Those eigenvalues will be shown to be estimators of the weights appearing in the asymptotic null distribution of the test. Hence, the method based on the eigenvalues is expected to provide a fine approximation of the distribution of the test. However, this approximation does not actually work well. Therefore, we further modify the method based on the eigenvalues, and the obtained method provides a better approximation.

The rest of this article is structured as follows. Section 2 introduces the framework of the two-sample test and defines the test statistics based on the MVD. In addition, the representation of test statistics based on the centered Gram matrices is described. Section 3.1 develops the asymptotics for the test by the MVD under \( H_0 \). The test by the MVD under \( H_1 \) is addressed in Section 3.2. Furthermore, the behavior of the test by the MVD under the local alternative hypothesis is clarified in Section 3.3. Section 3.4 describes the estimation of the weights that appear in the asymptotic null distribution obtained in Section 3.1. Section 4 examines the implementation of the MVD test with a Gaussian kernel in the Hilbert space \( \mathcal{H} = \mathbb{R}^d \). Section 4.1 introduces the modification of the approximate distribution given in Section 3.4. Section 4.2 reports the results of simulations for the type I error and the power of MVD and MMD tests. Section 5 presents the results of applications to real datasets, including high-dimension low-sample-size data. Conclusions are given in Section 6. All proofs of theoretical results are provided in the supplemental material.

### 2. Maximum variance discrepancy

Let \( \mathcal{H} \) be a separable Hilbert space and \( (\mathcal{H}, \mathcal{A}) \) be a measurable space. Let \( \langle \cdot, \cdot \rangle_{\mathcal{H}} \) be the inner product of \( \mathcal{H} \) and \( \| \cdot \|_{\mathcal{H}} = \sqrt{\langle \cdot, \cdot \rangle_{\mathcal{H}}} \) be the associated norm. Let \( X_1, \ldots, X_n \in \mathcal{H} \) and \( Y_1, \ldots, Y_m \in \mathcal{H} \) denote a sample of independent and identically distributed (i.i.d.) random variables drawn from unknown distributions \( P \) and \( Q \), respectively. Our goal is to test whether the unknown distributions \( P \) and \( Q \) are equal.

Let us define the null hypothesis \( H_0 : P = Q \) and the alternative hypothesis \( H_1 : P \neq Q \). Following Gretton et al. (2007), the gap between two distributions \( P \) and \( Q \) on \( \mathcal{H} \) is measured by:

\[
\text{MMD}(P, Q) = \sup_{f \in \mathcal{F}} \left( \mathbb{E}_{X \sim P}[f(X)] - \mathbb{E}_{Y \sim Q}[f(Y)] \right),
\]

where \( \mathcal{F} \) is a class of real-valued functions on \( \mathcal{H} \). Regardless of \( \mathcal{F} \), \( \text{MMD}(P, Q) \) always defines a pseudo-metric on the space of probability distributions. Let \( \mathcal{F} \) be the unit ball of a reproducing kernel Hilbert space \( H(k) \) associated with a characteristic kernel \( k : \mathcal{H} \times \mathcal{H} \to \mathbb{R} \) (see Aronszajn 1950; Fukumizu et al. 2009 for details) and assume that \( \mathbb{E}_{X \sim P}[\sqrt{k(X, X)}] < \infty \) and \( \mathbb{E}_{Y \sim Q}[\sqrt{k(Y, Y)}] < \infty \). Then, the MMD in \( H(k) \) is defined as the distance between \( P \) and \( Q \) as follows:

\[
\text{MMD}(P, Q) = \sup_{\|f\|_{H(k)} \leq 1} \langle f, \mu_k(P) - \mu_k(Q) \rangle_{H(k)} = \| \mu_k(P) - \mu_k(Q) \|_{H(k)},
\]

where \( \mu_k(P) = \mathbb{E}_{X \sim P}[k(\cdot, X)] \) and \( \mu_k(Q) = \mathbb{E}_{Y \sim Q}[k(\cdot, Y)] \) are called kernel mean embeddings of \( P \) and \( Q \), respectively, in \( H(k) \) (see Gretton et al. 2007 for details). The MMD focuses on the difference between distributions \( P \) and \( Q \) depending on the difference between the means of \( k(\cdot, X) \) and \( k(\cdot, Y) \) in \( H(k) \). The motivation for this research is to focus
on the difference between the distributions $P$ and $Q$ due to the difference between those variances in $H(k)$, based on a similar idea as the MMD. Assume $\mathbb{E}_{X \sim P}[k(X, X)] < \infty$ and $\mathbb{E}_{Y \sim Q}[k(Y, Y)] < \infty$, then the variance $V_{X \sim P}[k(\cdot, X)] : H(k) \to H(k)$ is defined by

$$V_{X \sim P}[k(\cdot, X)] = \mathbb{E}_{X \sim P}[(k(\cdot, X) - \mu_k(P))^{\otimes 2}] \in H(k)^{\otimes 2}.$$ 

Here, for any $h, h' \in H(k)$, the tensor product $h \otimes h'$ is defined as the operator $H(k) \to H(k), x \mapsto \langle h, x \rangle_{H(k)} h, h' \otimes h$ is defined as $h^{\otimes 2} = h \otimes h$, and $H(k)^{\otimes 2} = H(k) \otimes H(k)$ (see Section II.4 in Reed and Simon 1981 for details). Let $V_{X \sim P}[k(\cdot, X)] = \Sigma_k(P)$ and $V_{Y \sim Q}[k(\cdot, Y)] = \Sigma_k(Q)$. Let $A$ be an operator in the unit ball of $H(k)^{\otimes 2}$. Then, we define the MVD in $H(k)$ as

$$\text{MVD}(P, Q) = \sup_{\|A\|_{H(k)^{\otimes 2}} \leq 1} \langle A, \Sigma_k(P) - \Sigma_k(Q) \rangle_{H(k)^{\otimes 2}} = \|\Sigma_k(P) - \Sigma_k(Q)\|_{H(k)^{\otimes 2}},$$

which can be seen as a discrepancy between distributions $P$ and $Q$. The $T^2 = \text{MVD}(P, Q)^2$ can be estimated by

$$\hat{T}_{n,m}^2 = \|\Sigma_k(\hat{P}) - \Sigma_k(\hat{Q})\|_{H(k)^{\otimes 2}}^2,$$

where

$$\Sigma_k(\hat{P}) = \frac{1}{n} \sum_{i=1}^n (k(\cdot, X_i) - \mu_k(\hat{P}))^{\otimes 2}, \quad \mu_k(\hat{P}) = \frac{1}{n} \sum_{i=1}^n k(\cdot, X_i)$$

and

$$\Sigma_k(\hat{Q}) = \frac{1}{m} \sum_{j=1}^m (k(\cdot, Y_j) - \mu_k(\hat{Q}))^{\otimes 2}, \quad \mu_k(\hat{Q}) = \frac{1}{m} \sum_{j=1}^m k(\cdot, Y_j).$$

Let the Gram matrices be $K_X = (k(X_i, X_j))_{1 \leq i, j \leq n}$, $K_Y = (k(Y_j, Y_t))_{1 \leq j, t \leq m}$, and $K_{XY} = (k(X_i, Y_j))_{1 \leq i \leq n}$; the centering matrix be $P_n = I_n - (1/n)1_n1_n^T$; and the centered Gram matrices be $\tilde{K}_X = P_nK_XP_n$, $\tilde{K}_Y = P_mKYPD$, and $\tilde{K}_{XY} = P_nK_{XY}P_m$, where $I_n$ is the $n \times n$ identity matrix and $1_n$ is an $n$-dimensional vector having all of its components equal to 1. This test statistic can be expanded as

$$\hat{T}_{n,m}^2 = \frac{1}{n^2} \text{tr}(\tilde{K}_X^2) - \frac{2}{nm} \text{tr}(\tilde{K}_{XY} \tilde{K}_{XY}^T) + \frac{1}{m^2} \text{tr}(\tilde{K}_Y^2).$$

We investigate the behaviors of $\text{MMD}(N(0_d, I_d), N(t_1, s_{I_d}))^2$ and $\text{MVD}(N(0_d, I_d), N(t_1, s_{I_d}))^2$ for the difference of $(t, s)$ from $(0, 1)$ when $\mathcal{H} = \mathbb{R}^d$, the kernel $k(\cdot, \cdot)$ is the Gaussian kernel:

$$k(x, y) = \exp \left(-\sigma \|x - y\|_{\mathbb{R}^d}^2 \right), \quad \sigma > 0,$$

where $0_d$ is an $d$-dimensional vector of all 0’s. Those behaviors are obtained by using Corollary 1 in the supplementary material.

A sensitive reaction to the difference of $(t, s)$ from $(0, 1)$ means that it can sensitively react to differences between distributions from $N(0, I_d)$.

More generally, kernel $k'(x, y) = \exp(C)k(x, y)$ based on a constant $C$ and a positive definite kernel $k(x, y)$ is also positive definite. Then, $\text{MMD}_k(P, Q)$ and $\text{MVD}_k(P, Q)$ calculated by the kernel $k'$ hold $\text{MMD}_k(P, Q)^2 = \exp(C)\text{MMD}_k(P, Q)^2$ and $\text{MVD}_k(P, Q)^2 = \exp(2C)\text{MVD}_k(P, Q)^2$ for any distributions $P$ and $Q$ using $\text{MMD}_k(P, Q)$ and $\text{MVD}_k(P, Q)$ calculated by the kernel $k$. 


The graph of $\text{MMD}_{k'}(P, Q)^2$ and $\text{MVD}_{k'}(P, Q)^2$ is displayed for each $t$ when $s = 1$ in Figure 1 and for each $s$ when $t = 0$ in Figure 2. The kernel $k$ is a Gaussian kernel in (4), and the parameters are $C = 0, 4, 10$, $d = 10$, and $\sigma = 0.1$ in both Figures 1 and 2. Figure 1 shows the $\text{MMD}_{k'}(P, Q)^2$ and $\text{MVD}_{k'}(P, Q)^2$ for the difference of the mean from the standard normal distribution. In Figure 1a, $\text{MMD}_{k'}(P, Q)^2$ is larger than $\text{MVD}_{k'}(P, Q)^2$, but in Figure 1a and c, where the value of $C$ is increased, $\text{MVD}_{k'}(P, Q)^2$ is larger than $\text{MMD}_{k'}(P, Q)^2$ for each $t$. In addition, Figure 2 shows the reaction of the MMD and MVD to the difference of the covariance matrix from the standard normal distribution, and $\text{MVD}_{k'}(P, Q)^2$ is larger than $\text{MMD}_{k'}(P, Q)^2$ for each $s$ when $C$ is large. This means that MVD is more sensitive to differences from the standard normal distribution than MMD for $k'$ with large $C$.

![Figure 1](image1.png)

**Figure 1.** The $\text{MMD}_{k'}(N(0_d, I_d), N(t_{1_d}, sI_d))^2$ (solid), and $\text{MVD}_{k'}(N(0_d, I_d), N(t_{1_d}, sI_d))^2$ (dashed) for each $t$: $s = 1, d = 10, \sigma = 0.1$, and (a) $C = 0$, (b) $C = 4$, and (c) $C = 10$.

![Figure 2](image2.png)

**Figure 2.** The $\text{MMD}_{k'}(N(0_d, I_d), N(t_{1_d}, sI_d))^2$ (solid), and $\text{MVD}_{k'}(N(0_d, I_d), N(t_{1_d}, sI_d))^2$ (dashed) for each $s$: $t = 0, d = 10, \sigma = 0.1$, and (a) $C = 0$, (b) $C = 4$, and (c) $C = 10$. 
We have just observed that MVD and MMD take different values as discrepancies between two normal distributions. However, note that this different behavior of MVD and MMD is not directly related to the power of two-sample test as described in the following sections.

3. Test statistic for two-sample problem

We consider a two-sample test based on $T^2$ for $H_0 : P = Q$ and $H_1 : P \neq Q$, and $\hat{T}^2_{n,m}$ is defined as a test statistic. If $\hat{T}^2_{n,m}$ is large, then the null hypothesis $H_0$ is rejected since $T^2$ is the difference between $P$ and $Q$. The condition to derive the asymptotic distribution of this test statistic is as follows:

**Condition.** $\mathbb{E}_X \sim p[k(X, X)^2] < \infty$, $\mathbb{E}_Y \sim q[k(Y, Y)^2] < \infty$ and $\lim_{n,m \to \infty} n/(n + m) \to \rho$, $0 < \rho < 1$.

3.1. Asymptotic null distribution

In this section, we derive an asymptotic distribution of $\hat{T}^2_{n,m}$ under $H_0$. In what follows, the symbol “$\overset{D}{\rightarrow}$” designates convergence in distribution.

**Theorem 3.1.** Suppose that Condition is satisfied. Then, under $H_0 : P = Q$, as $n, m \to \infty$,

$$(n + m)\hat{T}^2_{n,m} \overset{D}{\rightarrow} \frac{1}{\rho(1 - \rho)} \sum_{\ell=1}^{\infty} \lambda_\ell Z^2_\ell,$$

where $Z_\ell \overset{i.i.d.}{\sim} N(0, 1)$ and $\lambda_\ell$ is an eigenvalue of $V_X \sim p[k(\cdot, X) - \mu_k(P)] \otimes^2$.

It is generally not easy to utilize such an asymptotic null distribution because it is an infinite sum and determining the weights included in the asymptotic distribution is itself a difficult problem. For practical purposes, a method to approximate the distribution of the test by $\hat{T}^2_{n,m}$ under $H_0$ is developed in **Section 4**. The method is based on the eigenvalues of the centered Gram matrices associated with the dataset in **Section 3.4**.

3.2. Asymptotic non null distribution

In this section, an asymptotic distribution of $\hat{T}^2_{n,m}$ under $H_1$ is investigated.

**Theorem 3.2.** Suppose that Condition is satisfied. Then, under $H_1 : P \neq Q$, as $n, m \to \infty$,

$$\sqrt{n + m}(\hat{T}^2_{n,m} - T^2) \overset{D}{\rightarrow} N(0, 4\rho^{-1}v^2_P + 4(1 - \rho)^{-1}v^2_Q),$$

where

$$v^2_P = V_X \sim p\left[\left(\Sigma_k(P) - \Sigma_k(Q), (k(\cdot, X) - \mu_k(P)) \otimes^2 - \Sigma_k(P)\right)_{H(k) \otimes^2}\right]$$

and

$$v^2_Q = V_Y \sim q\left[\left(\Sigma_k(P) - \Sigma_k(Q), (k(\cdot, Y) - \mu_k(Q)) \otimes^2 - \Sigma_k(Q)\right)_{H(k) \otimes^2}\right].$$
Remark 1. We see by Theorem 3.2 that
\[
\frac{\sqrt{n + m}(\hat{T}_{n,m}^2 - T^2)}{v} \rightarrow N(0, 1),
\]
where \( v = \sqrt{4\rho^{-1}v_P^2 + 4(1-\rho)^{-1}v_Q^2} \). Thus, we can evaluate the power of the test by \((n + m)\hat{T}_{n,m}^2\) as
\[
\Pr((n + m)\hat{T}_{n,m}^2 \geq t_\alpha | H_1) = \Pr((n + m)(\hat{T}_{n,m}^2 - T^2) \geq t_\alpha - (n + m)T^2 | H_1)
\]
\[
\approx 1 - \Phi \left( \frac{t_\alpha - \sqrt{n + mT^2}}{\sqrt{n + mv}} \right) \rightarrow 1
\]
as \( n, m \rightarrow \infty \), where \( t_\alpha \) is the \((1 - \alpha)\)-quantile of the distribution of \((n + m)\hat{T}_{n,m}^2\) under \( H_0 \), and \( \Phi \) is the distribution function of the standard normal distribution. Therefore, this test is consistent.

3.3. Asymptotic distribution under contiguous alternatives

In this section, we develop an asymptotic distribution of \( \hat{T}_{n,m}^2 \) under a sequence of local alternative distributions \( Q_{nm} = (1 - 1/\sqrt{n + m})P + (1/\sqrt{n + m})Q \) for \( P \neq Q \).

**Theorem 3.3.** Let \( X_1, \ldots, X_n \overset{i.i.d.}{\sim} P \) and \( Y_1, \ldots, Y_m \overset{i.i.d.}{\sim} Q_{nm} \). Suppose that Condition is satisfied. Let \( h : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R} \) be a kernel defined as
\[
h(x, y) = (k(\cdot, x) - \mu_k(P))^\otimes2 - \Sigma_k(P), (k(\cdot, y) - \mu_k(P))^\otimes2 - \Sigma_k(P)|_{H(k)^\otimes2}, \ x, y \in \mathcal{H}
\]
and
\[
h(\cdot, x) = (k(\cdot, x) - \mu_k(P))^\otimes2 - \Sigma_k(P) \in H(k)^\otimes2
\]
and let \( S_k : L_2(\mathcal{H}, P) \rightarrow L_2(\mathcal{H}, P) \) be a self-adjoint operator defined as
\[
S_k g(x) = \int_{\mathcal{H}} h(x, y) g(y) dP(y), \ g \in L_2(\mathcal{H}, P)
\]
(see Sections VI.1, VI.3, and VI.6 in Reed and Simon 1981 for details). Then, as \( n, m \rightarrow \infty \)
\[
(n + m) \left\| \Sigma_k(\hat{P}) - \Sigma_k(\hat{Q}_{nm}) \right\|_{H(k)^\otimes2}^2 \overset{D}{\rightarrow} \frac{1}{\rho(1 - \rho)} \sum_{\ell=1}^\infty \theta_\ell W_\ell^2,
\]
where \( W_\ell \sim N(\sqrt{\rho(1 - \rho) \cdot \xi_\ell(P, Q)/\theta_\ell}, 1) \), \( W_\ell \perp W_{\ell'} (\ell \neq \ell') \),
\[
\xi_\ell(P, Q) = \int_{\mathcal{H}} (\Sigma_k(Q) - \Sigma_k(P) + (\mu_k(Q) - \mu_k(P))^\otimes2, h(\cdot, y)|_{H(k)^\otimes2} \Psi_\ell(y) dP(y),
\]
and \( \theta_\ell \) and \( \Psi_\ell \) are, respectively, the eigenvalue of \( S_k \) and the eigenfunction corresponding to \( \theta_\ell \).

The following proposition claims that the eigenvalues \( \theta_\ell \) appearing in Theorem 3.3 are the same as the eigenvalues \( \lambda_\ell \) appearing in Theorem 3.1:
Proposition 3.4. The eigenvalues of $V_{X \sim P}[(k(\cdot, X) - \mu_k(P))^\otimes 2]$ in Theorem 3.1 and $S_k$ in (6) of Theorem 3.3 are the same.

From Theorems 3.1 and 3.3 and Proposition 3.4, it can be seen that the local power of the test by $(n + m)\varepsilon_{n,m}^2$ is dominated by the noncentrality parameters. It follows that

$$
\zeta_\ell(P, Q) = \int_{H} \{E_{Y \sim Q}[h(\cdot, y)], h(\cdot, x)\} \Psi_\ell(y) dP(y) = \lambda_\ell E_{Y \sim Q}[\Psi_\ell(Y)]
$$

by which we obtain

$$
E \left[ \frac{1}{\rho(1 - \rho)} \sum_{\ell=1}^\infty \theta_\ell W_\ell^2 \right] = \frac{1}{\rho(1 - \rho)} \sum_{\ell=1}^\infty \lambda_\ell \left( 1 + \rho(1 - \rho) \cdot \frac{\zeta_\ell(P, Q)^2}{\lambda_\ell^2} \right) = \frac{1}{\rho(1 - \rho)} \sum_{\ell=1}^\infty \lambda_\ell \left( 1 + \rho(1 - \rho) \{E_{Y \sim Q}[\Psi_\ell(Y)]\}^2 \right).
$$

In addition, from Theorem 1 in Minh, Niyogi, and Yao (2006), local power of MVD test can be seen as a function of

$$
LPF_{MVD}(P, Q) = \sum_{\ell=1}^\infty \lambda_\ell \{E_{Y \sim Q}[\Psi_\ell(Y)]\}^2
$$

$$
= \sum_{\ell=1}^\infty \lambda_\ell E_{Y, Y' \sim Q}[h(Y, Y')]
$$

$$
= \left\| \Sigma_k(Q) - \Sigma_k(P) + (\mu_k(Q) - \mu_k(P))^{\otimes 2} \right\|_{H(k)^{\otimes 2}}^2.
$$

Hence, the local power of MVD test can be expressed by not only the difference between $\Sigma_k(Q)$ and $\Sigma_k(P)$ but also that between $\mu_k(Q)$ and $\mu_k(P)$. We can seen from Gretton et al. (2012) that local power of MMD test can be expressed as a function of $LPF_{MMD}(P, Q) = \left\| \mu_k(Q) - \mu_k(P) \right\|_{H(k)}^2$.

For comparison, the asymptotic variances of MVD and MMD tests under contiguous alternatives are denoted as $AV_{MVD}(P)$ and $AV_{MMD}(P)$, respectively. Further, we shall assume that, for $a > 0$, sets

$$
L_1(P, a) = \{Q \mid \left\| \mu_k(Q) - \mu_k(P) \right\|_{H(k)}^2 < a\}
$$

and

$$
L_2(P, a) = \{Q \mid \left\| \Sigma_k(Q) - \Sigma_k(P) \right\|_{H(k)^{\otimes 2}} > \left\| \mu_k(Q) - \mu_k(P) \right\|_{H(k)}^2 + \sqrt{a}\}
$$

are both non-empty. Then, we have for $Q \in L_1(P, a) \cap L_2(P, a)$ that

$$
LPF_{MVD}(P, Q) \geq \left\| \Sigma_k(Q) - \Sigma_k(P) \right\|_{H(k)^{\otimes 2}}^2 + \left\| \mu_k(Q) - \mu_k(P) \right\|_{H(k)}^4 - 2 \left\| \Sigma_k(Q) - \Sigma_k(P) \right\|_{H(k)^{\otimes 2}} \left\| \mu_k(Q) - \mu_k(P) \right\|_{H(k)}^2 > a > LPF_{MMD}(P, Q).
$$

Hence, for $P \in \{P \mid AV_{MVD}(P) < AV_{MMD}(P)\}$, MVD test tends to have locally more powerful than MMD test for $Q \in L_1(P, a) \cap L_2(P, a)$. 
3.4. Null distribution estimates using Gram matrix spectrum

The asymptotic null distribution was obtained in Theorem 3.1, but it is difficult to derive its weights. The following theorem shows that this weight can be estimated using the estimator of \( V[(k(\cdot, X) - \mu_k(P))^\otimes 2] \).

**Theorem 3.5.** Assume that \( \mathbb{E}_{X \sim P}[k(X, X)^2] < \infty \). Let \( \{\lambda_\ell\}_{\ell=1}^\infty \) and \( \{\hat{\lambda}_\ell^{(n)}\}_{\ell=1}^\infty \) be the eigenvalues of \( Y \) and \( \hat{Y}^{(n)} \), respectively, where

\[
Y = V \left[ (k(\cdot, X) - \mu_k(P))^\otimes 2 \right] \quad \text{and} \quad \hat{Y}^{(n)} = \frac{1}{n} \sum_{i=1}^n \left\{ (k(\cdot, X_i) - \mu_k(P))^{\otimes 2} - \Sigma_k(P) \right\}^{\otimes 2}.
\]

Then, as \( n \to \infty \)

\[
\sum_{\ell=1}^\infty \hat{\lambda}_\ell^{(n)} Z_\ell^2 \xrightarrow{\mathcal{D}} \sum_{\ell=1}^\infty \lambda_\ell Z_\ell^2,
\]

where \( Z_\ell \sim N(0, 1) \).

In addition, Proposition 3.6 claims the eigenvalues of \( \hat{Y}^{(n)} \) and the Gram matrix are the same.

**Proposition 3.6.** The \( n \times n \) Gram matrix \( H = (H_{ij})_{1 \leq i,j \leq n} \) is defined as

\[
H_{ij} = \left( (k(\cdot, X_i) - \mu_k(P))^{\otimes 2} - \Sigma_k(P), (k(\cdot, X_j) - \mu_k(P))^{\otimes 2} - \Sigma_k(P) \right)_{H(k)^{\otimes 2}}.
\]

Then, the eigenvalues of \( \hat{Y}^{(n)} \) and \( H/n \) are the same.

**Remark 2.** By Proposition 3.6, the critical value can be obtained by calculating \( 1/\{\rho(1 - \rho)\} \sum_{\ell=1}^{n-1} \hat{\lambda}_\ell^{(n)} z_\ell^2 \) using the eigenvalue \( \hat{\lambda}_\ell^{(n)} \), \( \ell \in \{1, \ldots, n - 1\} \) of \( H/n \). In addition, the matrix \( H \) is expressed as

\[
H = P_n(\tilde{K}_X \odot \tilde{K}_X)P_n,
\]

where \( \odot \) is the Hadamard product. The \( n \times n \) Gram matrix \( K_X \) is a positive definite, but \( H \) has eigenvalue 0 since \( H_{11} = 0_n \).

4. Implementation

This section proposes corrections to the asymptotic distribution for both the MVD and MMD tests, and describes the results of simulations of the type-I error and power for the modifications. The MMD test is a two-sample test for \( H_0 \) and \( H_1 \) using the test statistic:

\[
\hat{\Delta}_{n,m}^2 = \left\| \frac{1}{n} \sum_{i=1}^n k(\cdot, X_i) - \frac{1}{m} \sum_{j=1}^m k(\cdot, Y_j) \right\|^2_{H(k)}.
\]

The \( \hat{\Delta}_{n,m}^2 \) of the asymptotic null distribution is the infinite sum of the weighted chi-square distribution, which is the same as \( \hat{T}_{n,m}^2 \) in (3). The approximate distribution can be obtained by estimating the eigenvalues by the centered Gram matrix (see Gretton et al. 2009 for details).
4.1. Approximation of the null distribution

In this section, we discuss methods to approximate the null distributions of the MVD and MMD tests. The asymptotic null distribution of the MVD test was obtained in Theorem 3.1 as an infinite sum of weighted chi-squared random variables with one degree of freedom, and according to Theorem 3.5, those weights \( \lambda_\ell \) (\( \ell \geq 1 \)) can be estimated by the eigenvalues \( \hat{\lambda}_\ell^{(n)} \) of the matrix \( \hat{H}/n \). Similar results were obtained for MMD by Gretton et al. (2009). However, this approximate distribution based on this estimated eigenvalue does not actually work well. In fact, by comparing the simulated exact null distribution with this approximate distribution based on estimated eigenvalues, it can be seen that variance of the approximate distribution is larger than that of the simulated exact null distribution. We modify the approximate distribution \( 1/(\rho(1-\rho)) \sum_{\ell=1}^{n-1} \hat{\lambda}_\ell^{(n)} Z_\ell^2 \) by obtaining the variance of this simulated exact null distribution. The variance of the exact null distribution \( V[(n+m)\hat{T}_{n,m}^2] \) is obtained as the following proposition:

**Proposition 4.1.** Assume that \( \mathbb{E}_{X \sim p}[k(X,X)^2] < \infty \). Then under \( H_0 \),

\[
V[(n+m)\hat{T}_{n,m}^2] = \frac{2(n+m)^4}{n^2 m^2} \| \gamma \|_{H(k) \otimes 4}^2 + O\left( \frac{1}{n} \right) + O\left( \frac{1}{m} \right).
\]

Proposition 4.1 leads to

\[
V[(n+m)\hat{T}_{n,m}^2] \approx \frac{(n+m)^4}{n^2 m^2} \cdot \frac{s^2 \ell^2}{(s+\ell)^4} V[(s+\ell)\hat{T}_{s,\ell}^2],
\]

with respect to \( V[(n+m)\hat{T}_{n,m}^2] \) and \( V[(s+\ell)\hat{T}_{s,\ell}^2] \), \( s, \ell \in \mathbb{N} \). If we can estimate the variance \( V[(s+\ell)\hat{T}_{s,\ell}^2] \) at \( s \) and \( \ell \) that is less than \( n \) and \( m \), respectively, we can estimate \( V[(n+m)\hat{T}_{n,m}^2] \) by using (8). In addition, the method of estimating \( V[(s+\ell)\hat{T}_{s,\ell}^2] \) by choosing \( (s, \ell) \) from \( (n, m) \) without replacement is known as subsampling.

The following proposition for MMD shows a similar result to MVD:

**Proposition 4.2.** Assume that \( \mathbb{E}_{X \sim p}[k(X,X)] < \infty \). Then, under \( H_0 \)

\[
V[(n+m)\hat{\Delta}_{n,m}^2] = \frac{2(n+m)^4}{n^2 m^2} \| \Sigma_k(P) \|_{H(k) \otimes 2}^2 + O\left( \frac{1}{n} \right) + O\left( \frac{1}{m} \right).
\]

4.1.1. Subsampling method

We used the subsampling method to estimate \( V[(s+\ell)\hat{T}_{s,\ell}^2] \) and \( V[(s+\ell)\hat{\Delta}_{s,\ell}^2] \) (see Section 2.2 in Politis et al. 1999 for details). In order to obtain two samples under the null hypothesis, we divide \( X_1, \ldots, X_n \) into \( X_1, \ldots, X_{n_1} \) and \( X_{n_1+1}, \ldots, X_n \). Then, we randomly select \( X_i^*(i), \ldots, X_k^*(i) \) and \( Y_i^*(i), \ldots, Y_k^*(i) \) from \( X_1, \ldots, X_{n_1} \) and \( X_{n_1+1}, \ldots, X_n \) without replacement, which repeat in each iteration \( i \in \{1, \ldots, I\} \). These randomly selected values generate the replicates of the test statistic

\[
\hat{T}_{s,\ell}^2(i) = \hat{T}_{s,\ell}^2(X_1^*(i), \ldots, X_k^*(i); Y_1^*(i), \ldots, Y_k^*(i))
\]
for iterations \( i \in \{1, \ldots, I \} \). The generated test statistics \( (s + \ell) \hat{T}_{s,\ell}^2(i) \) in \( I \) iterations estimate \( V((s + \ell) \hat{T}_{s,\ell}^2) \) by calculating the unbiased sample variance:

\[
V((s + \ell) \hat{T}_{s,\ell}^2)_{\text{sub}} = \frac{1}{I-1} \sum_{j=1}^{I} \left\{ (s + \ell) \hat{T}_{s,\ell}^2(j) - (s + \ell) \bar{T}_{s,\ell}^2 \right\}^2,
\]

where \( \bar{T}_{s,\ell}^2 = (1/I) \sum_{i=1}^{I} \hat{T}_{s,\ell}^2(i) \). According to (8), \( V((n + m) \hat{T}_{n,m}^2) \) is estimated by

\[
V((n + m) \hat{T}_{n,m}^2)_{\text{sub}} = \frac{(n + m)^4}{n^2m^2} \frac{s^2\ell^2}{(s + \ell)^4} V((s + \ell) \bar{T}_{s,\ell}^2)_{\text{sub}}.
\]

We also estimate \( V((n + m) \hat{\Delta}_{n,m}^2) \) by using

\[
V((n + m) \hat{\Delta}_{n,m}^2)_{\text{sub}} = \frac{(n + m)^4}{n^2m^2} \frac{s^2\ell^2}{(s + \ell)^4} V((s + \ell) \bar{\Delta}_{s,\ell}^2)_{\text{sub}}
\]

from Proposition 4.2, where

\[
\hat{\Delta}_{s,\ell}^2(i) = \hat{\Delta}_{s,\ell}^2(X_1^s(i), \ldots, X_s^s(i); Y_1^s(i), \ldots, Y_s^s(i))
\]

for \( i \in \{1, \ldots, I\} \),

\[
V((s + \ell) \hat{\Delta}_{s,\ell}^2)_{\text{sub}} = \frac{1}{I-1} \sum_{j=1}^{I} \left\{ (s + \ell) \hat{\Delta}_{s,\ell}^2(j) - (s + \ell) \bar{\Delta}_{s,\ell}^2 \right\}^2
\]

and \( \bar{\Delta}_{s,\ell}^2 = (1/I) \sum_{i=1}^{I} \hat{\Delta}_{s,\ell}^2(i) \).

The columns of \((n + m) \hat{T}_{n,m}^2\) in Table 1 and \((n + m) \hat{\Delta}_{n,m}^2\) in Table 2 are variances of \((n + m) \hat{T}_{n,m}^2\) and \((n + m) \hat{\Delta}_{n,m}^2\), which are estimated by a simulation of 10,000 iterations with \( X_1, \ldots, X_n \sim N(0_d, I_d) \) and \( Y_1, \ldots, Y_m \sim N(0_d, I_d) \) for each \( \sigma, d \) and \( (n, m) \). The subsampling variances \( V((n + m) \hat{T}_{n,m}^2)_{\text{sub}} \) in (9) and \( V((n + m) \hat{\Delta}_{n,m}^2)_{\text{sub}} \) in (10) with \( X_1, \ldots, X_n \sim N(0_d, I_d) \) for each \( \sigma, d \) and \( (n, m) \). The columns of \((n + m) \hat{T}_{n,m}^2\) under \( P = Q = N(0_d, I_d) \) and \( V((n + m) \hat{T}_{n,m}^2)_{\text{sub}} \): \( I = 1,000, n_1 = n/2, \) and \( X_1, \ldots, X_n \sim N(0_d, I_d) \).

### Table 1. The variance of \((n + m) \hat{T}_{n,m}^2\) under \( P = Q = N(0_d, I_d) \) and \( V((n + m) \hat{T}_{n,m}^2)_{\text{sub}} \):

| \( \sigma \) | \( d \) | \( (n, m) \) | \( (n + m) \hat{T}_{n,m}^2 \) | \( (n/4, n/4) \) | \( (n/6, n/6) \) | \( (n/8, n/8) \) |
|---|---|---|---|---|---|---|
| \( d^{-3/4} \) | 5 | (200,200) | 0.06880 | 0.04341 | 0.05168 | 0.04902 |
| \( d^{-3/4} \) | 5 | (500,500) | 0.06881 | 0.03821 | 0.04897 | 0.04921 |
| \( d^{-7/8} \) | 5 | (200,200) | 0.07254 | 0.04246 | 0.05138 | 0.05798 |
| \( d^{-7/8} \) | 5 | (500,500) | 0.07188 | 0.04052 | 0.05500 | 0.05593 |
| \( d^{-3/4} \) | 10 | (200,200) | 0.00850 | 0.00602 | 0.00812 | 0.00898 |
| \( d^{-3/4} \) | 10 | (500,500) | 0.00845 | 0.00674 | 0.00751 | 0.00753 |
| \( d^{-7/8} \) | 10 | (200,200) | 0.01280 | 0.00937 | 0.01224 | 0.01377 |
| \( d^{-7/8} \) | 10 | (500,500) | 0.01270 | 0.01032 | 0.01251 | 0.01255 |
| \( d^{-3/4} \) | 20 | (200,200) | 0.00049 | 0.00048 | 0.00070 | 0.00094 |
| \( d^{-3/4} \) | 20 | (500,500) | 0.00043 | 0.00031 | 0.00046 | 0.00060 |
| \( d^{-7/8} \) | 20 | (200,200) | 0.00166 | 0.00152 | 0.00261 | 0.00330 |
| \( d^{-7/8} \) | 20 | (500,500) | 0.00147 | 0.00122 | 0.00165 | 0.00204 |

slope of the line

| (500,500) | 1 | 1.63621 | 1.35769 | 1.29601 |
| (both) | 1 | 1.76057 | 1.33845 | 1.3232 |
| (both) | 1 | 1.69348 | 1.34798 | 1.30928 |
Table 2. The variance of \((n + m)\hat{\Delta}_{n,m}^2\) under \(P = Q = N(0, d, I_d)\) and \(V[(n + m)\hat{\Delta}_{n,m}^2]\) sub: \(I = 1,000, n_1 = n/2,\) and \(X_1, \ldots, X_n \sim N(0, d, I_d)\).

| \(\sigma\) | \(d\) | \((n, m)\) | \((n + m)\hat{\Delta}_{n,m}^2\) sub | Subsampling \((s, \ell)\) |
|------------|--------|-------------|-----------------|-----------------|
| \(-3/4\)   | 5      | (200,200)   | 0.57100         | 0.47044         |
| \(-3/4\)   | 5      | (500,500)   | 0.66068         | 0.54518         |
| \(-7/8\)   | 5      | (200,200)   | 0.65987         | 0.51867         |
| \(-7/8\)   | 5      | (500,500)   | 0.75903         | 0.63563         |
| \(-3/4\)   | 10     | (200,200)   | 0.16205         | 0.09213         |
| \(-3/4\)   | 10     | (500,500)   | 0.16279         | 0.16106         |
| \(-7/8\)   | 10     | (200,200)   | 0.25656         | 0.14104         |
| \(-7/8\)   | 10     | (500,500)   | 0.25836         | 0.24670         |
| \(-3/4\)   | 20     | (200,200)   | 0.02757         | 0.02135         |
| \(-3/4\)   | 20     | (500,500)   | 0.02784         | 0.02255         |
| \(-7/8\)   | 20     | (200,200)   | 0.07642         | 0.07404         |
| \(-7/8\)   | 20     | (500,500)   | 0.07856         | 0.05714         |

The variance of Table 2.

Remark 3. Since the subsampling variance \(V[(s + \ell)\hat{T}_{s,\ell}^2]\) sub is an unbiased sample variance of \(I\) times, we get

\[
\mathbb{E}[V[(s + \ell)\hat{T}_{s,\ell}^2]\text{sub}] = \frac{1}{I-1} \sum_{j=1}^I \mathbb{E}\left[\left((s + \ell)\hat{T}_{s,\ell}^2(j) - (s + \ell)\overline{T}_{s,\ell}^2\right)^2\right]
\]

\[
= \frac{1}{I-1} \sum_{j=1}^I \mathbb{E}\left[\left((s + \ell)\hat{T}_{s,\ell}^2(j) - \mathbb{E}[(s + \ell)\hat{T}_{s,\ell}^2] + \mathbb{E}[(s + \ell)\hat{T}_{s,\ell}^2] - (s + \ell)\overline{T}_{s,\ell}^2\right)^2\right]
\]
Figure 3. The results of linear regression of the simulated variance of \((n + m) \hat{\Delta}^2_{n,m}\) and the subsampling variance \(V[\{(n + m) \hat{\Delta}^2_{n,m} \}_{\text{sub}}]\) and the results for the MMD. (a) MVD \((n, m) = (200, 200), (s, \ell) = (50, 50)\). (b) MVD \((n, m) = (500, 500), (s, \ell) = (125, 125)\). (c) MMD \((n, m) = (200, 200), (s, \ell) = (50, 50)\). (d) MMD \((n, m) = (500, 500), (s, \ell) = (125, 125)\).

\[
\begin{align*}
&= \frac{1}{I-1} \sum_{j=1}^{I} \left[ \mathbb{E}\left\{ (s + \ell) \hat{T}^2_{s,\ell}(j) - \mathbb{E}(s + \ell) \hat{T}^2_{s,\ell} \right\}^2 \right] + \mathbb{E}\left[ \left\{ \mathbb{E}(s + \ell) \hat{T}^2_{s,\ell} - (s + \ell) \overline{T}^2_{s,\ell} \right\}^2 \right] \\
&\quad + 2 \mathbb{E}\left[ \left\{ (s + \ell) \hat{T}^2_{s,\ell}(j) - \mathbb{E}(s + \ell) \hat{T}^2_{s,\ell} \right\} \left\{ \mathbb{E}(s + \ell) \hat{T}^2_{s,\ell} - (s + \ell) \overline{T}^2_{s,\ell} \right\} \right] \\
&= \frac{1}{I-1} \sum_{j=1}^{I} \left\{ V[(s + \ell) \hat{T}^2_{s,\ell}] + \frac{1}{I^2} \sum_{i,t=1}^{I} \operatorname{Cov}\left((s + \ell) \hat{T}^2_{s,\ell}(i), (s + \ell) \hat{T}^2_{s,\ell}(t)\right) \right\} \\
&\quad - \frac{2}{I} \sum_{i=1}^{I} \operatorname{Cov}\left((s + \ell) \hat{T}^2_{s,\ell}(j), (s + \ell) \hat{T}^2_{s,\ell}(i)\right) 
\end{align*}
\]
In particular, a Monte Carlo simulation is performed to observe the type-I error and the regression coefficient in the row “slope of the line” in Tables 1 and 2. Figure 4 shows that this approximation method can be similarly discussed for the MMD test using $\hat{\theta}_{\text{exact}}$ and $\hat{\theta}_{\text{approx}}$, as seen in Tables 1 and 2, we use positive $\tau > 0$ to estimate $V[(n + m)\hat{\theta}_{\text{exact}}]$ by $(1 + \tau) V[(n + m)\hat{\theta}_{\text{approx}}]$. This underestimation is the same for the MMD test and $V[(n + m)\hat{\theta}_{\text{approx}}]$ is estimated by $(1 + \tau) V[(n + m)\hat{\theta}_{\text{approx}}]$ with positive $\tau > 0$. Our approximation of the null distribution is based on a modification of the large variance of $1/\{(\rho - 1)\} \sum_{\ell=1}^{\infty} \hat{\lambda}_{\ell}^{(n)} Z_{\ell}^{2}$ to $(1 + \tau) V[(n + m)\hat{\theta}_{\text{approx}}]$. The method aims to approximate the exact null distribution by using

$$W'_n = \frac{\xi_n}{\{(\rho - 1)\}} \sum_{\ell=1}^{n-1} \hat{\lambda}_{\ell}^{(n)} Z_{\ell}^{2} + c_n. \quad (11)$$

The parameters $\xi_n$ and $c_n$ are determined so that the means of $W'_n$ and $1/\{(\rho - 1)\} \sum_{\ell=1}^{n-1} \hat{\lambda}_{\ell}^{(n)}$ are equal and the variance of $W'_n$ is equal to $(1 + \tau) V[(n + m)\hat{\theta}_{\text{approx}}]$, which can be established by

$$\mathbb{E}[W'_n] = \frac{1}{\rho - 1} \sum_{\ell=1}^{\infty} \hat{\lambda}_{\ell}^{(n)}$$

and

$$V[W'_n] = (1 + \tau) V[(n + m)\hat{\theta}_{\text{approx}}].$$

This approximation method can be similarly discussed for the MMD test using $(1 + \tau) V[(n + m)\hat{\theta}_{\text{approx}}]$. In this article, the parameter $\tau > 0$ is determined using the value of “slope of the line” in Tables 1 and 2. Figure 4 shows that this $W'_n$ can approximate the simulated exact distribution better than $1/\{(\rho - 1)\} \sum_{\ell=1}^{n-1} \hat{\lambda}_{\ell}^{(n)} Z_{\ell}^{2}$. The algorithm for the MMD test can be obtained by changing $H$ and $\hat{\theta}$ in Algorithm to $\hat{\theta}_{\text{approx}}$. 4.2. Simulations

In this section, we investigate the performance of $(n + m)\hat{\theta}_{\text{approx}}$ under a specific null hypothesis and specific alternative hypotheses when $H = \mathbb{R}^d$ and $k(\cdot, \cdot)$ is the Gaussian kernel in (4). In particular, a Monte Carlo simulation is performed to observe the type-I error and the power of MVD and MMD tests. Two cases are implemented: a uniform distribution $Q_1$ and
an exponential distribution $Q_2$, with $P = N(0,1)$, all of which have means and variances 0 and 1, respectively. The critical values are determined based on $W_n'$ in Section 4.1.2 from a normal distribution. The type-I error of $(n + m)\hat{T}_{n,m}^2$ can be obtained by counting the number of times $(n + m)\hat{T}_{n,m}^2$ exceeds the critical value in 1,000 iterations under the null hypothesis. Next, the estimated power of $(n + m)\hat{T}_{n,m}^2$ is similarly obtained by counting the number of times $(n + m)\hat{T}_{n,m}^2$ exceeds the critical value under each alternative distribution in 1,000 iterations. We execute the above for $(n, m) = (200, 200)$ and $(500, 500)$ and $d = 5, 10$, and 20. It is known that the selection of the value of $\sigma$ involved in the Gaussian kernel affects the performance. We utilize $\sigma$ depending on dimension $d$. The significance level is $\alpha = 0.05$. The critical values are determined on the basis of $W_n'$ in Section 4.1.2 from a normal distribution. The type-I error and estimated power can be obtained by counting how many times $(n + m)\hat{T}_{n,m}^2$ exceeds the critical values in 1,000 iterations under $P = Q$ and $P \neq Q$. With $n_1 = n/2$ and $(s, \ell) = (n/8, n/8)$, $\tau_{\text{MVD}}$ for MVD is $0.30928$ and $\tau_{\text{MMD}}$ for MMD is $0.11643$ by subtracting 1 from the value for “slope of the line” for “both” in Tables 1 and 2. The following can be seen from Table 3:

**Algorithm 1** Calculation of critical value for the MVD test.

**Input:** $X_1, \ldots, X_n, Y_1, \ldots, Y_m \in \mathcal{H}$, $k : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ (kernel), $0 < \alpha < 1$ (significance level) and $(s, \ell) \in \{1, \ldots, [n/2]\}$, $\tau > 0$ (parameters). For example, $\tau$ is selected as shown in the values of “slope of the line” in Tables 1 and 2.

1. Compute the eigenvalues $\hat{\lambda}_\ell^{(n)}$ of $H$ in (7) and obtain $1/(\rho(1 - \rho)) \sum_{\ell=1}^{n-1} \hat{\lambda}_\ell^{(n)} Z_\ell^2(j)$ by random element $Z_1^{(j)}, \ldots, Z_{n-1}^{(j)} \sim N(0,1)$, $j \in \{1, \ldots, J\}$.

2. (a) Obtain copies $(s + \ell)\hat{T}_{s,\ell}^2(i)$, $i \in \{1, \ldots, I\}$ of $(s + \ell)\hat{T}_{s,\ell}^2$ under $H_0$ by the subsampling method.

(b) Compute subsampling variance $V[(n + m)\hat{T}_{n,m,\text{sub}}^2]_{\text{sub}}$ from $(s + \ell)\hat{T}_{s,\ell}^2(i)$.

3. Compute $(n + m)\hat{T}_{n,m,\text{sub}}^2$ in (11) by $1/(\rho(1 - \rho)) \sum_{\ell=1}^{n-1} \hat{\lambda}_\ell^{(n)} Z_\ell^2(j)$ and $V[(n + m)\hat{T}_{n,m,\text{sub}}^2]$.

**Output:** We obtain the critical value $t_\alpha(W_n')$ as the $(J(1 - \alpha)\text{-th})$ from the top sorted in ascending order.
Table 3. Type-I error and power of test by $(n + m)\hat{T}_{n,m}^2$ for each sample size and each parameter $\sigma$.

| $\sigma$ | $d$ | $(n, m)$ | MVD | $Q_1$ | $Q_2$ | Type-I error | $Q_1$ | $Q_2$ |
|----------|-----|----------|-----|-------|-------|---------------|-------|-------|
| $d^{-3/4}$ | 5   | (200,200) | 0.060 | 0.797 | 1     | 0.047         | 0.401 | 1     |
| $d^{-3/4}$ | 5   | (500,500) | 0.072 | 1     | 1     | 0.063         | 0.877 | 1     |
| $d^{-7/8}$ | 5   | (200,200) | 0.056 | 0.728 | 1     | 0.052         | 0.305 | 1     |
| $d^{-7/8}$ | 5   | (500,500) | 0.067 | 0.999 | 1     | 0.053         | 0.735 | 1     |
| $d^{-3/4}$ | 10  | (200,200) | 0.073 | 0.612 | 1     | 0.086         | 0.342 | 1     |
| $d^{-3/4}$ | 10  | (500,500) | 0.047 | 0.991 | 1     | 0.040         | 0.630 | 1     |
| $d^{-7/8}$ | 10  | (200,200) | 0.054 | 0.482 | 1     | 0.086         | 0.235 | 1     |
| $d^{-7/8}$ | 10  | (500,500) | 0.034 | 0.955 | 1     | 0.044         | 0.363 | 1     |
| $d^{-3/4}$ | 20  | (200,200) | 0.279 | 0.816 | 1     | 0.082         | 0.239 | 1     |
| $d^{-3/4}$ | 20  | (500,500) | 0.099 | 0.948 | 1     | 0.068         | 0.477 | 1     |
| $d^{-7/8}$ | 20  | (200,200) | 0.060 | 0.332 | 1     | 0.047         | 0.113 | 0.989 |
| $d^{-7/8}$ | 20  | (500,500) | 0.034 | 0.728 | 1     | 0.069         | 0.240 | 1     |

- Table 3 shows that the probabilities of type-I error at $d = 5$ and 10 are near the significance level of $\alpha = 0.05$ for both the MVD and MMD.
- The probability of type-I error at $d = 20$ exceeds the significance level of $\alpha = 0.05$ for the MVD, but decreases as $(n, m)$ increases.
- It can be seen that the critical value by $W'_n$ of the MVD tends to be estimated as less than that point of the null distribution.
- In hypothesis $P = Q_1$, it can be seen that the MVD has a higher power than the MMD.
- It can be seen that the MVD and MMD have higher powers for hypothesis $P = Q_2$ than hypothesis $P = Q_1$ and it is difficult to distinguish between the normal distribution and the uniform distribution by the MVD and MMD for a Gaussian kernel.
- Note that the critical value changes depending on the distribution of the null hypothesis.

Tables 1–3, which are the simulation results in Section 4, are all for the simple case of $n = m$ and $s = \ell$. These simulations can also be implemented with $n \neq m$, but the $n \neq m$ case is not investigated in this article because of the increasing number of combinations of $\sigma, d, n, m$. In practical use, it is desirable to use the larger of the two samples to construct $W'_n$.

5. Application to real datasets

The MVD test was applied to some real datasets. The significance level was $\alpha = 0.05$ and the critical value $t_{0.05}(H)$ was obtained through 10,000 iterations of $1/(\rho(1 - \rho)) \sum_{\ell=1}^{n-1} \lambda_{\ell}^{(n)} Z_{\ell}^2$ based on the eigenvalues of the matrix $H/n$. We also calculated the critical value $t_{0.05}(W'_n)$ of the approximate distribution $W'_n$ according to Algorithm in Section 4.1.2. The $t_{0.05}(\tilde{K}_X)$ calculates the critical value for the MMD test from the distribution obtained based on Theorem 1 in Gretton et al. (2009) through 10,000 iterations. See the supplementary material for results applied to datasets other than Colon.

5.1. Colon data

The Colon dataset contains gene expression data from DNA microarray experiments of colon tissue samples with $d = 2,000$ and $n = 62$ (see Alon et al. 1999 for details). Among the 62
Table 4. Values of \( (n + m) \hat{T}_{n,m}^2 \) and critical values for normal \( (n_1 = 11, s, \ell = 6) \) and tumor \( (n_1 = 20, s, \ell = 10) \), with \( \tau_{MVD} = 0.69348 \).

| \( \sigma \) | tumor vs. normal | \( t_{0.05}(W_n^s) \) | \( t_{0.05}(H) \) | tumor | \( t_{0.05}(W_n^s) \) | \( t_{0.05}(H) \) |
|------------|-----------------|-----------------|-----------------|------|-----------------|-----------------|
| \( d^{-3/4} \) | 3.867 | 3.536 | 5.280 | 3.728 | 5.050 |
| \( d^{-7/8} \) | 2.291 | 2.097 | 2.907 | 2.258 | 2.846 |
| \( d^{-1} \) | 0.684 | 0.660 | 0.879 | 0.757 | 0.906 |

Table 5. Values of \( (n + m) \hat{T}_{n,m}^2 \) and critical values for normal \( (n_1 = 11, s, \ell = 6) \) and tumor \( (n_1 = 20, s, \ell = 10) \), with \( \tau_{MMD} = 0.21990 \).

| \( \sigma \) | tumor vs. normal | \( t_{0.05}(W_n^s) \) | \( t_{0.05}(H) \) | tumor | \( t_{0.05}(W_n^s) \) | \( t_{0.05}(H) \) |
|------------|-----------------|-----------------|-----------------|------|-----------------|-----------------|
| \( d^{-3/4} \) | 6.695 | 4.456 | 6.201 | 4.618 | 5.713 |
| \( d^{-7/8} \) | 8.787 | 3.974 | 4.827 | 3.945 | 4.491 |
| \( d^{-1} \) | 6.282 | 2.439 | 2.754 | 2.412 | 2.634 |

Table 6. Values of \( (n + m) \hat{T}_{n,m}^2 \) and critical values for tumor 1 \( (n_1 = 10, s, \ell = 5) \) and tumor 2 \( (n_1 = 10, s, \ell = 5) \), with \( \tau = 0.69348 \).

| \( \sigma \) | tumor 1 vs. tumor 2 | \( t_{0.05}(W_n^s) \) | \( t_{0.05}(H) \) | tumor 1 | \( t_{0.05}(W_n^s) \) | \( t_{0.05}(H) \) |
|------------|-----------------|-----------------|-----------------|------|-----------------|-----------------|
| \( d^{-3/4} \) | 3.379 | 3.180 | 4.596 | 3.245 | 4.942 |
| \( d^{-7/8} \) | 1.858 | 1.915 | 2.502 | 2.085 | 2.875 |
| \( d^{-1} \) | 0.558 | 0.629 | 0.800 | 0.727 | 0.921 |

Table 7. Values of \( (n + m) \hat{T}_{n,m}^2 \) and critical values for tumor 1 \( (n_1 = 10, s, \ell = 5) \) and tumor 2 \( (n_1 = 10, s, \ell = 5) \), with \( \tau = 0.21990 \).

| \( \sigma \) | tumor 1 vs. tumor 2 | \( t_{0.05}(W_n^s) \) | \( t_{0.05}(H) \) | tumor 1 | \( t_{0.05}(W_n^s) \) | \( t_{0.05}(H) \) |
|------------|-----------------|-----------------|-----------------|------|-----------------|-----------------|
| \( d^{-3/4} \) | 4.627 | 4.064 | 5.621 | 3.961 | 5.782 |
| \( d^{-7/8} \) | 4.123 | 3.305 | 4.206 | 3.426 | 4.551 |
| \( d^{-1} \) | 2.453 | 1.942 | 2.377 | 2.102 | 2.656 |

samples, 40 are tumor tissues and 22 are normal tissues. Tables 4 and 5 show the results of the MVD and MMD tests for \( P = \text{tumor} \) and \( Q = \text{normal} \). The “tumor vs. normal” column shows the values of \( (n + m) \hat{T}_{n,m}^2 \) and \( (n + m) \hat{T}_{n,m}^2 \) for \( P = \text{tumor} \) and \( Q = \text{normal} \). The “normal” and “tumor” columns show, \( t_{0.05}(W_n^s) \) and \( t_{0.05}(H) \) calculated from, respectively, the normal tissues and tumor tissues datasets.

For the MVD, \( (n + m) \hat{T}_{n,m}^2 \) does not exceed \( t_{0.05}(H) \), but \( (n + m) \hat{T}_{n,m}^2 \) exceeds \( t_{0.05}(W_n^s) \) by modifying the approximate distribution. By contrast, for the MMD, \( (n + m) \hat{T}_{n,m}^2 \) exceeds both \( t_{0.05}(H) \) and \( t_{0.05}(W_n^s) \) without modifying the approximate distribution.

Next, tumor (sample size = 40) was divided into \( P = \text{tumor 1} (n = 20) \) and \( Q = \text{tumor 2} (m = 20) \), and two-sample tests by the MVD and MMD were applied. The results are shown in Tables 6 and 7, with the values for \( (n + m) \hat{T}_{n,m}^2 \) and \( (n + m) \hat{T}_{n,m}^2 \) in the column “tumor 1 vs. tumor 2”. In Table 6, when \( \sigma = d^{-3/4} \), \( (n + m) \hat{T}_{n,m}^2 \) exceeds \( t_{0.05}(W_n^s) \), but in other cases \( (n + m) \hat{T}_{n,m}^2 \) does not exceed \( t_{0.05}(W_n^s) \) and \( (n + m) \hat{T}_{n,m}^2 \) does not exceed \( t_{0.05}(H) \) and \( t_{0.05}(W_n^s) \).
Table 7 shows that, for all $\sigma$, $(n + m)\hat{\Delta}^2_{n,m}$ exceeds $t_{0.05}(W'_n)$ for the MMD test, but $(n + m)\hat{\Delta}^2_{n,m}$ does not exceed $t_{0.05}(H)$.

6. Conclusion

We defined a Maximum Variance Discrepancy (MVD) with a similar idea to the Maximum Mean Discrepancy (MMD) in Section 2. We derived the asymptotic null distribution for the MVD test in Section 3.1. This was the infinite sum of the weighted chi-square distributions. In Section 3.2, we derived an asymptotic non null distribution for the MVD test, which was a normal distribution. The asymptotic normality of the test under the alternative hypothesis showed that the two-sample test by the MVD has consistency. Furthermore, we developed an asymptotic distribution for the test under a sequence of local alternatives in Section 3.3. This was the infinite sum of weighted non central chi-squared distributions. We constructed an estimator of asymptotic null distributed weights based on the Gram matrix in Section 3.4. The approximate distribution of the null distribution by these estimated weights does not work well, so we modified it in Section 4.1. In the simulation of the power reported, we found that the power of MVD test was larger than that of MMD. We confirmed in Section 5 that the two-sample test by the MVD works for real data-sets.

Acknowledgments

The authors would like to express their deepest thanks to the Editor-in-Chief for his helpful suggestions. I would like to thank the two anonymous referees, whose suggestions greatly improved the paper. I am also grateful to Professor Kanta Naito for a lot of helpful discussions and also for valuable comments. The research of the author is supported in part by KAKENHI 22K17860.

References

Alon, U., N. Barkai, D. A. Notterman, K. Gish, S. Ybarra, D. Mack, and A. J. Levine. 1999. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proceedings of the National Academy of Sciences of the United States of America* 96 (12):6745–50. 10.1073/pnas.96.12.6745. 10359783

Aronszajn, N. 1950. Theory of reproducing kernels. *Transactions of the American Mathematical Society* 68 (3):337–404. 10.1090/S0002-9947-1950-0051437-7.

Boente, G., D. Rodriguez, and M. Sued. 2018. Testing equality between several populations covariance operators. *Annals of the Institute of Statistical Mathematics* 70 (4):919–50. 10.1007/s10463-017-0613-1.

Fukumizu, K., A. Gretton, X. Sun, and B. Schölkopf, 2009. Kernel measures of conditional dependence. Paper presented Advances in Neural Information Processing Systems 20 - Proceedings of the 2007 Conference, Cambridge, MA, MIT Press, 489–96.

Gretton, A., K. M. Borgwardt, M. Rasch, B. Schölkopf, and A. J. Smola. 2007. A kernel method for the two-sample-problem. In *Advances in neural information processing systems*, ed. by B. Schölkopf, J. C. Platt, and T. Hoffman, Vol. 19, 513–20. Cambridge, MA: MIT Press.

Gretton, A., K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola. 2012. A kernel two-sample test. *Journal of Machine Learning Research* 13 (25):723–73.

Gretton, A., K. Fukumizu, Z. Harchaoui, and B. Sriperumbudur. 2009. A fast, consistent kernel two-sample test. *Advances in Neural Information Processing Systems*, Vol. 22, 673–81.

Kim, Y.-J., C. M. Nam, and J. Kim. 2013. Nonparametric two-sample tests of longitudinal data with termination events. *Communications in Statistics- Theory and Methods* 42 (6):1056–69. 10.1080/03610926.2011.592255.
Minh, H. Q., P. Niyogi, and Y. Yao. 2006. Mercer’s theorem, feature maps, and smoothing, In *Proceedings of the 19th annual conference on learning theory*, 154–68. Berlin, Heidelberg: Springer-Verlag.

Politis, D., D. Wolf, J. Romano, M. Wolf, P. Bickel, P. Diggle, and S. Fienberg. 1999. Subsampling. New York: Springer.

Reed, M., and B. Simon. 1981. *Functional analysis*. Elsevier Science, Academic Press.

Thi Thien Trang, B., J.-M. Loubes, L. Risser, and P. Balaresque. 2021. Distribution regression model with a reproducing Kernel Hilbert space approach. *Communications in Statistics- Theory and Methods* 50 (9):1955–77. 10.1080/03610926.2019.1658782.

Zhang, C., H. Peng, and J.-T. Zhang. 2010. Two samples tests for functional data. *Communications in Statistics- Theory and Methods* 39 (4):559–78. 10.1080/03610920902755839.