A SPECTRAL TWO SAMPLE TEST

A Thesis
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Yangshuang Xu

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Department:
Statistics

May 2021

Fargo, North Dakota

NORTH DAKOTA STATE UNIVERSITY

Graduate School

Title

A SPECTRAL TWO SAMPLE TEST

By

Yangshuang Xu

The supervisory committee certifies that this thesis complies with North Dakota State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Prof. Mingao Yuan

Chair

Prof. Bong-jin Choi

Prof. Changhui Yan

Approved:

June 14, 2021

Date

Prof. Rhonda Magel

Department Chair

ABSTRACT

In statistics, two-sample tests are commonly desired under many topics. It's crucial to differentiate populations as a prerequisite for developing further analysis. A number of statistical tests have been developed for this hypothesis testing problem. In this thesis, we propose a novel method to perform two-sample test. Specifically, we use the two samples to form a matrix and adopt the largest eigen-value as the test statistic. This test statistic followed the tracy-widom law as the limiting distribution under the null hypothesis. We evaluate the performance of the proposed method by extensive simulation study and real data application. The type I error is consistently and asymptotically controlled to nominal level. Our test manifests competitive power and prevailing calculation cost compared with several well-known two-sample test methods. The real data application also shows the advantage of the proposed method.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDIX TABLES

1. INTRODUCTION

1.1. Background

In the mid-twentieth century, Neyman and Pearson proposed a study to determine if two observed samples arise from the same population (Neyman and Pearson 1933). Several years later, Frank Wilcoxon also worked on a similar problem (Kruskal 1957). However, the popularity of solving such problem has not diminished as time goes by. We want to test if two samples are from the same population (null hypothesis) or not from the same population (alternative hypothesis). Many papers proposed their own methods for such hypothesis testing. These methods have a consistent name, called the "two sample test". Such test can be applied in diverse areas. In clinical research, we want to compare the effectiveness between treatments. In machine learning, one of the interests is to evaluate generative models (Arjovsky et al. 2017).

In general, there are two main categories of two-sample tests: parametric and non-parametric tests. Parametric tests include the well-known Student's t-test (Student 1908). Such tests usually rely on several presumptions (e.g., normality of sample distribution). These tests are able to calculate the power in a finite number of operations (closed-form). Nonetheless, parametric tests may lose the ability to differentiate between samples when they are not distributed as assumed. Non-parametric tests don't have a presumption for distribution. In other words, these tests can be performed for arbitrary distributions. However, non-parametric tests need excessive time for the resulting power because of the cross-validation or bootstrap scheme performed. A classical non-parametric test that has been applied to deep learning research is the kernel two-sample test, which calculates the empirical maximum mean discrepancy (MMD) between two populations as the test statistic. MMD test highly relies on the selection of characteristic kernels, which are not easily determined accurately at earlier stages of research (Liu et al. 2015). Besides, unsuitable choice of characteristic kernels could possibly result in lower power and higher computation cost.

In this thesis, we developed a novel two-sample test, which is named as the spectral two sample test. Specifically, we construct a symmetric matrix using the two samples and center and scale it. Then we calculate the largest eigen-value of the transformed matrix as the test statistic, which converges in distribution to the Tracy-Widom law under the null hypothesis (Tracy and

Widom 1993). The proposed method has computational $O(n\sqrt{n} + m\sqrt{m})$, where n, m are the sample sizes of the two sample. We evaluate the performance of the test statistic and compare it with well-known existing methods in terms of computational time and power. We also apply it to a real data. Our method has competitive power with faster computational time.

1.2. Two sample test and some well-known methods

Suppose we have independent and identically distributed samples $X_1, \dots, X_n \sim f$ and $Y_1, \dots, Y_m \sim g$. The two sample test problem is to test the following hypotheses:

$$\mathcal{H}_0 : f = g, \qquad \mathcal{H}_1 : f \neq g. \tag{1.1}$$

Under $\mathcal{H}_0$, the distributions of the two populations are equal. Under $\mathcal{H}_1$, the distributions of the two populations are not equal. The two sample test problem is one of the popular problems in statistics and machine learning. Many methods are available in the literature. Here we will introduce 4 well-known and widely applied two-sample tests and compare these tests with ours in the simulation chapter.

1.2.1. Kolmogorov-Smirnov test

Kolmogorov-Smirnov test (KS test) is a classical non-parametric two-sample test that has been used in many areas. This test calculates the largest distance ($D_{n,m}$) between the empirical distribution function (e.d.f.) from the respective sample. The KS test statistic is defined as (Hassani and Silva 2015)

$$D_{n,m} = \sup_{x,y} |F_{1,n}(x) - F_{2,m}(y)|, \tag{1.2}$$

where $F_{1,n}(x)$ and $F_{2,m}(y)$ are the e.d.f. of the respective sample. $\mathcal{H}_0$ is rejected at level of α if

$$D_{n,m} > c(\alpha)\sqrt{\frac{n+m}{n}}. \tag{1.3}$$

The computational cost is $O(m + n)$. Before testing $\mathcal{H}_1$, we need to judge if KS test is suitable for both distributions, for example, KS test can not be applied to discrete distributions, such as the Poisson distribution (Hassani and Silva 2015). Sometimes, it also achieves low empirical power under specific continuous distribution (e.g., F-distribution).

1.2.2. Maximum Mean Discrepancy test

Maximum Mean Discrepancy test (MMD test) is also a non-parametric two sample test. The test statistic is defined as the largest difference of expectation over functions in the reproducing kernel Hilbert space (RKHS). The authors proposed a distribution-free two sample test based on the asymptotic distribution of the unbiased statistic, The test statistic is defined as (Gretton et al. 2012)

$$\mathrm{MMD}_u^2 = \frac{1}{m(m-1)}\sum_{i=1}^{m}\sum_{j\neq i}^{m} k(x_i, x_j) + \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} k(y_i, y_j) - \frac{2}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} k(x_i, x_j), \quad (1.4)$$

where $k(x_i, x_j)$ represents a symmetric kernel function.

The asymptotic distribution is characterized by the following theorem (Gretton et al. 2012).

Theorem 1. *Let t=m+n, then* $\lim_{m,n\to\infty} m/t \to \rho_x$ *and* $lim_{m,n\to\infty} n/t \to \rho_y := (1-\rho_x)$ *for* $0 < \rho_x < 1$. *Under* $\mathcal{H}_0$, MMD_u^2, *in equation (1.4), converges to*

$$t\mathrm{MMD}_u^2 \to \sum_{l=1}^{\infty} \lambda_l \left[(\rho_x^{-1/2} a_l - \rho_y^{-1/2} b_l)^2 - (\rho_x \rho_y)^{-1} \right], \quad (1.5)$$

where a_l *and* b_l *are infinite sequences of independent Gaussian random variables, and they both follow Normal (0,1),* λ_i *is the eigenvalue of* $\int_{\mathcal{X}} k(x, x')\Psi_i(x)dp(x) = \lambda_i \Psi_i(x')$.

This test takes $O((m+n)^2)$ computation cost. To minimize the computation time, the authors proposed another test statistic when sample sizes are sufficient. The empirical estimator of the linear statistic is characterized in the following Lemma (Gretton et al. 2012):

Lemma 1. *Suppose* $m = n$, *define* $m_2 = [m/2]$, *and* $h(z_i, z_j) := k(x_i, x_j) + k(y_i, y_j) - k(x_i, y_j) - k(x_j, y_i)$, *the estimator*

$$\mathrm{MMD}_l^2 := \frac{1}{m_2}\sum_{i=1}^{m_2} h((x_{2i-1}, y_{2i-1}), (x_{2i}, y_{2i})) \quad (1.6)$$

can be computed in linear time, and it is also an unbiased estimator.

This test takes a computational cost of $O(m+n)$. The empirical power correspondingly decreases as the trade-off for computational speed under $\mathcal{H}_1$.

1.2.3. Anderson-Darling test

Anderson-Darling test (AD test) is first designed to test if one sample follows a specific distribution (Anderson and Darling 1952), for example, Weibull distribution. Later, Pettitt generalized the AD test to a non-parametric two sample test. Suppose $Z = Z_1, \ldots, Z_k$, which is the combined and ordered observations from $X_{(n)}$ and $Y_{(m)}$, and $k = m+n$. The test statistic is defined as (Pettitt 1976):

$$\text{AD} = \frac{1}{mn} \sum_{i=1}^{k-1} \frac{(kc_i - mi)^2}{i(k-i)}, \tag{1.7}$$

where $1 \leq i \leq k$, c_i = number of elements in $C_i = \{x : x \in X \text{ and } x \leq z_i\}$. The corresponding critical value can be found in Anderson-Darling table. AD test has good empirical power under $\mathcal{H}_1$. However, the computational cost is $O((m+n)^2)$, which is costly.

1.2.4. Mann-Whitney U test

Mann-Whitney U (MWU) test, proposed by Mann and Whitney, is an alternative to the two sample student's t-test. As a non-parametric test, MWU test is also distribution-free. The test statistic can be computed by the following steps (Mann and Whitney 1947):

(1) Rank the combined data points from smallest to largest (or largest to smallest).

(2) Calculate the sum of ranks in the first group and call it R_1. R_2 is similarly defined for the second group. Then, compute $U_1 = R_1 - n(n+1)/2$ and $U_2 = R_2 - m(m+1)/2$.

(3) Define $U = \min(U_1, U_2)$ as the test statistic.

$\mathcal{H}_1$ is rejected when test statistic is greater than the critical value, which depends on both sample sizes and can be found in MWU table. This test achieves comparable empirical power to the KS test, but it requires higher computation cost than the KS test. Besides, this test can possibly suggest low empirical power when f and g have similar mean.

1.3. Proposed method

Before introducing our two sample test, we will first discuss the theoretical basis behind our methodology.

1.3.1. Theoretical basis

Bickel and Sarkar 2016 propose a method to determine how many clusters existed in a network. The problem is formulated as a hypothesis testing. The null hypothesis is that there's only one cluster in the network. The alternative hypothesis is that there's more than one clusters in the network. The test statistic is calculated in the following steps (Bickel and Sarkar 2016):

(1) Create a centered and scaled adjacency matrix (A) of the network.

(2) Normalize matrix A to $\tilde{A}$.

(3) Calculate test statistic $\tilde{\theta}$ defined below

$$\tilde{\theta} = r^{\frac{2}{3}}(\lambda_1(\tilde{A}) - 2). \tag{1.8}$$

(4) Under $\mathcal{H}_0$, $\tilde{\theta} \to TW_1$. Reject null hypothesis if $\tilde{\theta}<\mathrm{TW}_{1,\alpha/2}$ or $\tilde{\theta}>\mathrm{TW}_{1,1-\alpha/2}$.

Here, r is the dimension of $\tilde{A}$, $\lambda_1(\tilde{A})$ is the largest eigenvalue of $\tilde{A}$ and TW_1 is the Tracy-Widom law with degree of freedom one. The TW_1 distribution is introduced by Tracy and Widom (Tracy and Widom 1993). They define TW_1 distribution as:

$$F_1(x) = exp\left[-\frac{1}{2}\int_x^\infty q(s)ds\right]\sqrt{exp\left(-\int_x^\infty (s-x)[q(s)]^2 ds\right)}, \tag{1.9}$$

where $q(s)$ represents the solution to Painlev´e type II equation. This limiting law is first proposed in paper Soshnikov 1999. Later, Lee and Yin generalize the adjacency matrix A to a random symmetric matrix (Lee and Yin 2014).

If there are more than one clusters exsited in the network, the network will be bipartitioned. Then repeat steps (1) to (4) until there's no more cluster can be found. Figure 1.1 shows a network after bipartitioning.

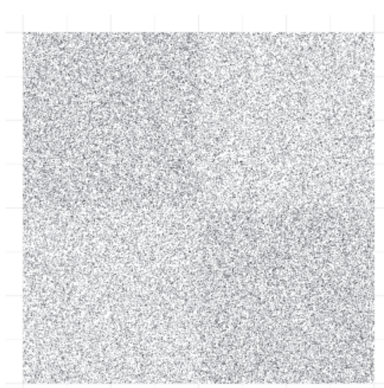

Figure 1.1. Network after bipartitioning.

In this case (Figure 1.1), the network has 2 clusters (Xu et al. 2017). The nodes from different clusters are partitioned by colors, they are dark gray (at top left and bottom right) and gray (other positions). Correspondingly, the adjacency matrix A partitions the elements (nodes in network) similarly (Bickel and Sarkar 2016).

1.3.2. Our methodology

The above method could be applied to two-sample test. We can consider the two samples as two clusters and form an adjacency matrix A. Then we calculate $\tilde{\theta}$ as the test statistic and compare the test statistic with corresponding critical values to conclude whether $\mathcal{H}_1$ is rejected at given nominal level. Specifically, our test statistic is calculated by the following steps:

(1) Create an upper triangular matrix (U). The elements in diagonal are 0. We fill the observations from the respective sample into U, row by row, as follows.

$$U_{ij} = \begin{cases} 0, & i = j \\ Y, & i \leq \frac{r}{2}, j \geq \frac{r}{2} \\ X, & others \end{cases}, \tag{1.10}$$

where r is the dimension of U and is defined as

$$r = \left\lfloor \frac{1 + \sqrt{1 + 8(n+m)}}{2} \right\rfloor. \tag{1.11}$$

Here r is computed by solving the equation $1 + 2 + \cdots + (r-1) = n + m$, which means total number of elements in U equals to the sum of sample sizes.

Here is an example for establishing U. Suppose we have a sample of 6 observations $(X_1, \ldots, X_6)$ from f and another sample of 9 observations $(Y_1, \ldots, Y_9)$ from g, then U is determined as:

$$U_{6\times 6} = \begin{pmatrix} 0 & X_1 & X_2 & Y_1 & Y_2 & Y_3 \\ & 0 & X_3 & Y_4 & Y_5 & Y_6 \\ & & 0 & Y_7 & Y_8 & Y_9 \\ & & & 0 & X_4 & X_5 \\ & & & & 0 & X_6 \\ & & & & & 0 \end{pmatrix}. \tag{1.12}$$

(2) Next, we will create matrix A by $A = U^T + U$. For the example above, the matrix A can be determined as:

$$A_{6\times 6} = \begin{pmatrix} 0 & \mathrm{X}_1 & \mathrm{X}_2 & \mathrm{Y}_1 & \mathrm{Y}_2 & \mathrm{Y}_3 \\ \mathrm{X}_1 & 0 & \mathrm{X}_3 & \mathrm{Y}_4 & \mathrm{Y}_5 & \mathrm{Y}_6 \\ \mathrm{X}_2 & \mathrm{X}_3 & 0 & \mathrm{Y}_7 & \mathrm{Y}_8 & \mathrm{Y}_9 \\ \mathrm{Y}_1 & \mathrm{Y}_4 & \mathrm{Y}_7 & 0 & \mathrm{X}_4 & \mathrm{X}_5 \\ \mathrm{Y}_2 & \mathrm{Y}_5 & \mathrm{Y}_8 & \mathrm{X}_4 & 0 & \mathrm{X}_6 \\ \mathrm{Y}_3 & \mathrm{Y}_6 & \mathrm{Y}_9 & \mathrm{X}_5 & \mathrm{X}_6 & 0 \end{pmatrix} \tag{1.13}$$

(3) We define the normalized matrix $\tilde{A}$ as:

$$\tilde{A} = \frac{A - \tilde{\mu}_{m+n}}{s_{m+n}}. \tag{1.14}$$

$$\tilde{A}_{r\times r} = \begin{pmatrix} \frac{-\tilde{\mu}_{m+n}}{s_{m+n}} & \frac{X_1-\tilde{\mu}_{m+n}}{s_{m+n}} & \frac{X_2-\tilde{\mu}_{m+n}}{s_{m+n}} & \cdots & \frac{Y_1-\tilde{\mu}_{m+n}}{s_{m+n}} & \cdots & \frac{Y_{r-\frac{n}{2}}-\tilde{\mu}_{m+n}}{s_{m+n}} \\ \frac{X_1-\tilde{\mu}_{m+n}}{s_{m+n}} & \frac{-\tilde{\mu}_{m+n}}{s_{m+n}} & \frac{X_{r/2}-\tilde{\mu}_{m+n}}{s_{m+n}} & \cdots & \frac{Y_{r-\frac{n}{2}+1}-\tilde{\mu}_{m+n}}{s_{m+n}} & \cdots & \frac{Y_{2r-n}-\tilde{\mu}_{m+n}}{s_{m+n}} \\ \frac{X_2-\tilde{\mu}_{m+n}}{s_{m+n}} & \frac{X_{r/2}-\tilde{\mu}_{m+n}}{s_{m+n}} & \frac{-\tilde{\mu}_{m+n}}{s_{m+n}} & \cdots & \frac{Y_{2r-n+1}-\tilde{\mu}_{m+n}}{s_{m+n}} & \cdots & \frac{Y_{3r-\frac{3}{2}n}-\tilde{\mu}_{m+n}}{s_{m+n}} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \frac{Y_1-\tilde{\mu}_{m+n}}{s_{m+n}} & \frac{Y_{r-\frac{n}{2}+1}-\tilde{\mu}_{m+n}}{s_{m+n}} & \frac{Y_{2r-n+1}-\tilde{\mu}_{m+n}}{s_{m+n}} & \cdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & \cdots & \frac{-\tilde{\mu}_{m+n}}{s_{m+n}} & \frac{X_n-\tilde{\mu}_{m+n}}{s_{m+n}} \\ \frac{Y_{r-\frac{n}{2}}-\tilde{\mu}_{m+n}}{s_{m+n}} & \frac{Y_{2r-n}-\tilde{\mu}_{m+n}}{s_{m+n}} & \frac{Y_{3r-\frac{3}{2}n}-\tilde{\mu}_{m+n}}{s_{m+n}} & \cdots & \cdots & \frac{X_n-\tilde{\mu}_{m+n}}{s_{m+n}} & \frac{-\tilde{\mu}_{m+n}}{s_{m+n}} \end{pmatrix}$$

.

In equation (1.14), $\tilde{\mu}_{m+n}=\frac{\sum_{i=1}^{k} Z_i}{k}$ is the mean of the combined sample, $s_{m+n} = \sqrt{\frac{\sum(z_i-\bar{z})}{k-1}}$ is the standard error of the combined sample. Here $Z = z_1, \ldots, z_k$ is the combined sample from X and Y, $k = m + n$.

(4) Define $\tilde{\theta}=r^{\frac{2}{3}}(\lambda_1(\tilde{A}) - 2)$ as our test statistic. As a two-sided test, $\mathcal{H}_1$ will be rejected if $\tilde{\theta}<\mathrm{TW}_{1,\alpha/2}$ or $\tilde{\theta}>\mathrm{TW}_{1,1-\alpha/2}$.

2. SIMULATION

R project is used throughout all phases of our methodology. Under $\mathcal{H}_0$ (f=g), we expect a consistently controlled type I error that closes to the nominal level. If this prerequisite is satisfied, we will compute the empirical power under $\mathcal{H}_1$ ($f \neq g$).

2.1. Simulation methods: Empirical type I error

The type I error represents the rejection of a true null hypothesis. Under $\mathcal{H}_0$, f and g are set as identical distribution. We generate random observations under respective distribution by R project. Then, we establish matrix U, A and $\tilde{A}$ by rule (1.10) to equation (1.14). $\mathcal{H}_0$ is rejected when $\tilde{\theta} \geq TW_{1,0.975} = 1.45$ or $\tilde{\theta} \leq TW_{1,0.025} = -3.51$. This procedure will be repeated for 500 times. The empirical type I error is calculated by the proportion of rejection times to the repetitions (500). We choose 8 commonly used distributions, including symmetric distribution (e.g., Normal distribution), skewed distribution (e.g., Gamma distribution) and discrete distribution (e.g., Poisson distribution). For each distribution, we compute the empirical type I errors under different parameters. The sample size in each distribution increases from 150 to 3750.

2.2. Simulation methods: Empirical power

If the empirical type I error is close to the nominal level, we will then focus on the empirical power. It is similarly calculated as empirical type I error, but now the observations are generated under $\mathcal{H}_1$ (when $f \neq g$). The empirical power is calculated by the proportion of rejection times to the repetitions (500). We will first show our empirical power when f and g have different parameters under same distribution (e.g, $f \sim$ Normal (1,1) and $g \sim$ Normal (0,2)). Next, we will evaluate the empirical power when f and g are different distributions (e.g., $f \sim$ Normal (0,1) and $g \sim$ t (2)). We mainly focus on large sample sizes (1000 to 5000) while calculating empirical power. The power and computation cost are compared between proposed method and other two sample tests.

2.3. Simulation result: Empirical type I error

Each of the following figure will show the empirical type I error under a given distribution. Table 2.1 will clarify the parameters in every commonly used distribution.

Table 2.1. Commonly used distribution with its parameters.

Distribution	Normal	student's t	Gamma	Binomial
Function	$\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{(x-\mu)}{\sigma}^2}$	$\frac{\Gamma(\frac{v+1}{2})}{\sqrt{v\pi}\Gamma(\frac{v}{2})})\left(1+\frac{x^2}{v}\right)^{-\frac{v+1}{2}}$	$\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-\frac{x}{\beta}}$	$\binom{m}{k}p^k q^{(m-k)}$
Parameters	μ, σ	$df(v)$	α, β	m, p
Distribution	Beta	Chi-square	Poisson	Weibull
Function	$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathbf{B}(\alpha,\beta)}$	$\frac{1}{2^{k/2}\Gamma(k/2)}x^{k/2-1}e^{-\frac{x}{2}}$	$\frac{\lambda^k e^{-\lambda}}{k!}$	$\frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\alpha-1}e^{-(\frac{x}{\alpha})^\beta}$
Parameters	α, β	$df(k)$	λ	α, β

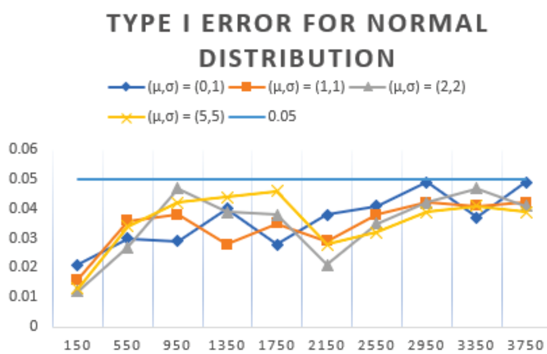

Figure 2.1. Empirical type I error under Normal distribution.

Figure 2.1 shows the empirical type I error under Normal distribution. The error gradually increases when sample size is less than 1750. Then it's stably controlled to 0.04 when sample size increases from 2150. Besides, we increase mean and variance from 0 to 5 and 1 to 5, respectively. Our method has consistent empirical type I error. On the whole, all the empirical type I errors are satisfactory.

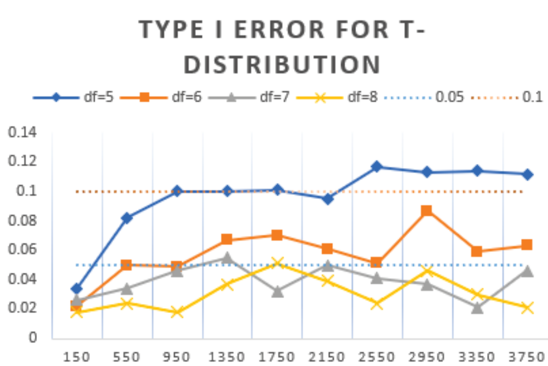

Figure 2.2. Empirical type I error under t-distribution.

Figure 2.2 provides the empirical type I error under student's t-distribution. As we know, t-distribution is similarly distributed as Normal distribution, but it has fatter tails. As a characteristic of t-distribution, when degree of freedom (df) increases, the distribution patterns will be more centered. From the outcome, our method has excellent empirical type I error when df=7 and 8 under all samples sizes. For smaller degree of freedom (df=6 and df=5), our methodology also has acceptable empirical type I error. A possible reason that results in higher error is that the sample sizes are not sufficient. Thus, we increase the sample size to 10,000. Then, the empirical type I error decreases to 0.06 and 0.07, respectively. As a result, the type I error under t-distribution is overall satisfactory. However, our method will require a more sufficient sample size to achieve better empirical error.

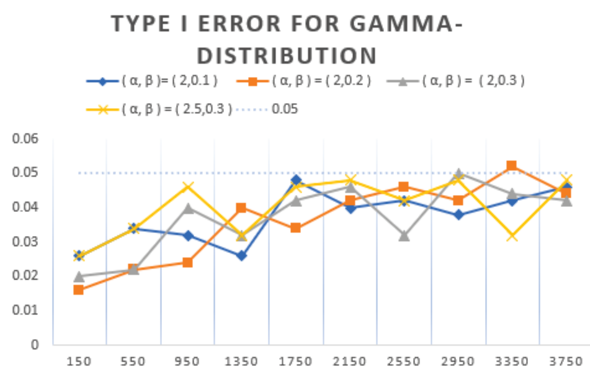

Figure 2.3. Empirical type I error under gamma distribution.

Figure 2.3 suggests the empirical type I error in Gamma distribution. Our method has good errors under Normal distribution and t-distribution, which are symmetric distributions. We want to evaluate the empirical error when the distribution is non-symmetric (skewed). We set α (shape parameter) from 2 to 2.5, both of them make Gamma distribution pattern significantly left-skewed (higher α results in more symmetric pattern). Consistently, our methodology has an average of 0.04 empirical type I error, which is very good.

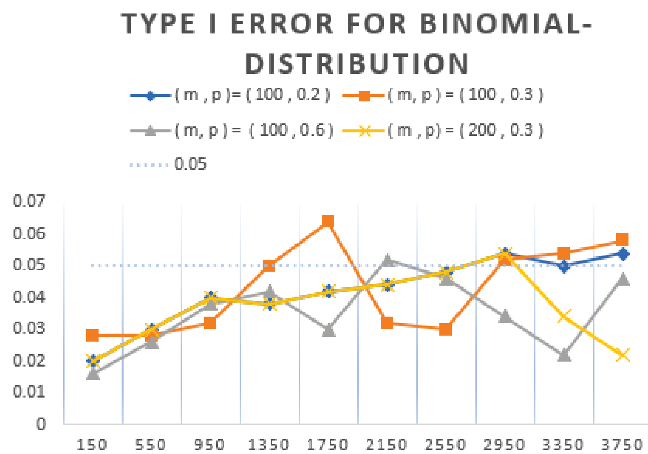

Figure 2.4. Empirical type I error under Binomial distribution.

Figure 2.4 presents the empirical type I error under Binomial distribution, which is a typical discrete distribution. The number of experiments (m) are set from 100 to 200, the distribution patterns are nearly symmetric under both m. For the first 3 groups, all m equal to 100, the empirical type I error is stably controlled when p is small. For groups (200,0.3) and (100,0.3), they have an identical mean that equals to 60 ($m \times p$). Our methodology does not provide very different empirical error. As a result, our methodology has good empirical error in Binomial distribution.

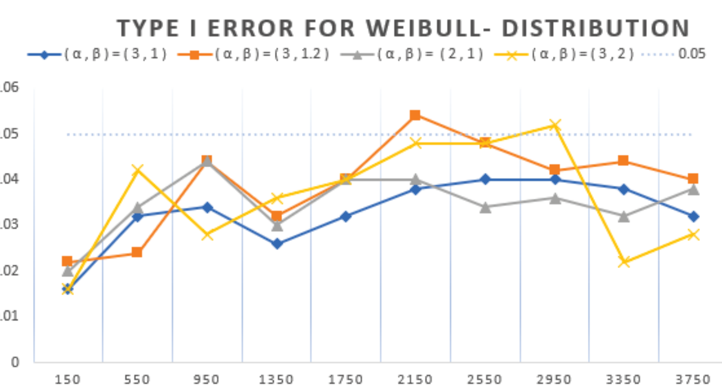

Figure 2.5. Empirical type I error under Weibull distribution.

Figure 2.5 shows the empirical type I error under Weibull-distribution. The pattern of Weibull-distribution is left-skewed when α=2, and it's nearly symmetric when α=3. The empirical type I error has an average of 0.03 under both situations, thus, we have well controlled empirical error under Wellbull-distribution.

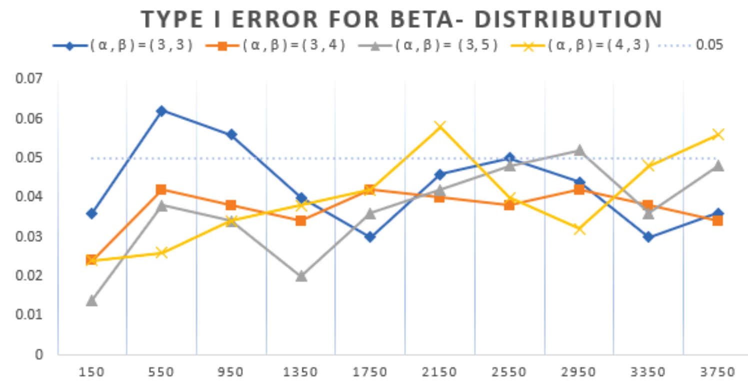

Figure 2.6. Empirical type I error under Beta distribution.

Figure 2.6 illustrates the empirical type I error under Beta distribution. This distribution can be left-skewed, symmetric or right-skewed. It's nearly symmetric under parameter (3,3), the type I error are controlled to 0.04 when sample sizes increase from 2150. Beta distribution are left-skewed under parameter (3,4) and (3,5), and right-skewed under parameter (4,3). On the whole, our empirical type I errors are consistently low.

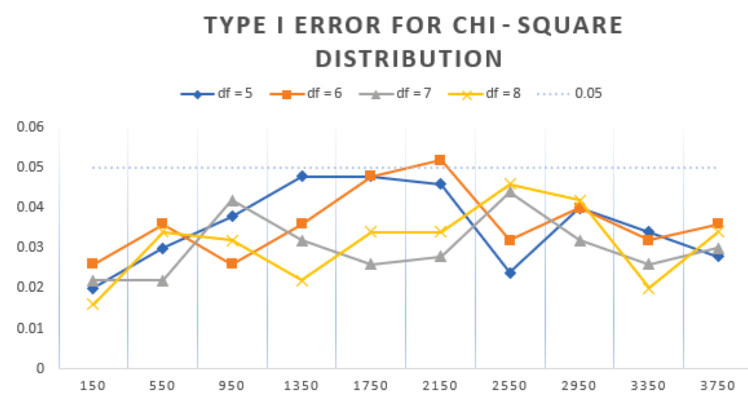

Figure 2.7. Empirical type I error under Chi-square distribution.

Figure 2.7 shows the empirical type I error under Chi-sqaure distribution. We increase the degree of freedom from 5 to 8, the patterns of chi-square will correspondingly be more centered. As a result, all the empirical type I error are close to an average of 0.03 under any degree of freedom.

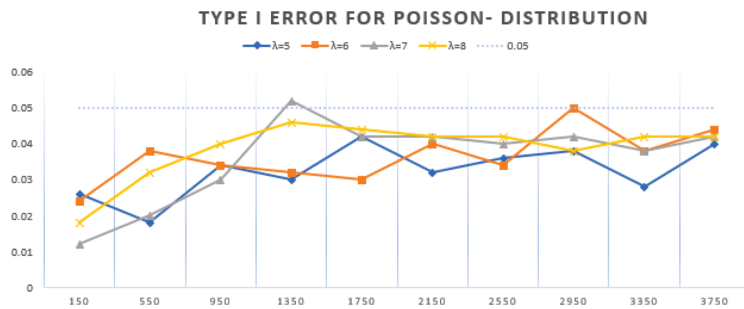

Figure 2.8. Empirical type I error under Poisson distribution.

Figure 2.8 presents the empirical type I error under Poisson disribution. This is also a discrete distribution. Differently, Figure 2.4 mainly discusses the type I error when the discrete distribution is nearly symmetric. In this case, we focus on a situation when the distribution is left-skewed. As the characteristic of Poisson distribution, the pattern will be more skewed under larger λ. From the outcome, all empirical type I error are close to an average of 0.04.

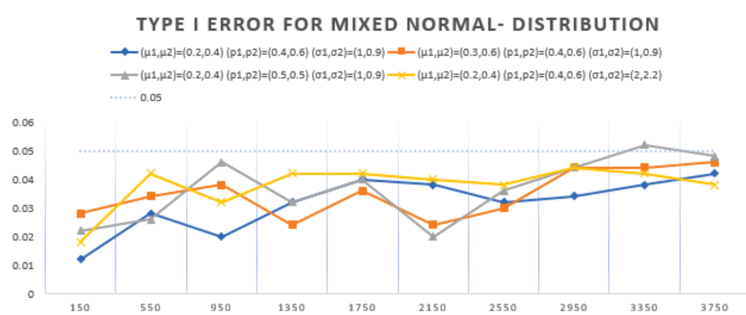

Figure 2.9. Empirical type I error under Mixed Normal distribution.

Figure 2.9 illustrates the empirical type I error under Mixed Normal distribution. The Mixed Normal distribution comes from two mixture components (Normal distribution) that are combined by weights p (or probabilities). If the components have different μ, the Mixed Normal distribution will have two crests. Though this is a complex distribution, our methodology still achieves excellent empirical type I error.

In conclusion, our method has satisfactory empirical error under commonly used distributions, no matter f and g are discrete or continuous, no matter they are symmetric or skewed. For most of the distributions discussed above, our empirical errors are controlled to an average of 0.04 even if sample sizes are not sufficient (less than 1750).

2.4. Simulation result: Empirical power

In section 2.4.1, we will show our empirical power when f and g are under same distribution with different parameters. In section 2.4.2, we will provide empirical power when f and g are under different distributions. Each following table will show the power and computation cost from both proposed method and commonly used tests.

2.4.1. Samples from same distribution with different parameters

In this part, we will develop the empirical power when f and g are under same distribution with different parameters.

Table 2.2. $f \sim$ F-distribution with $(df1, df2) = (2, 5)$ and $g \sim$ F-distribution with $(df1, df2) = (2, 8)$.

n	$\lambda_1(\tilde{A})$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	Anderson-Darling (time elapsed(s))	MWU (time elapsed(s))
1500	0.962 (8.62)	0.020 (3.03)	0.292 (2.17)	0.962 (1927.14)	0.346 (9.01)
2000	0.970 (11.02)	0.000 (3.45)	0.230 (2.80)	1.000 (3749.28)	0.416 (11.57)
2500	0.990 (15.50)	0.020 (3.99)	0.264 (3.08)	1.000 (8202.25)	0.500 (14.47)
3000	1.000 (19.38)	0.000 (4.04)	0.306 (3.68)	1.000 (14092.02)	0.612 (17.74)
3500	0.998 (25.17)	0.000 (4.39)	0.334 (4.24)	1.000 (20175.98)	0.700 (20.03)
4000	1.000 (29.29)	0.000 (5.9)	0.358 (5.96)	1.000 (26602.84)	0.714 (22.8)
4500	1.000 (35.94)	0.000 (6.2)	0.342 (5.20)	1.000 (34496.27)	0.764 (25.49)
5000	1.000 (40.92)	0.020 (5.95)	0.370 (5.72)	1.000 (41014.28)	0.872 (26.75)

Table 2.2 shows the empirical power when f and g are F-distribution with different parameters. Based on this table, our methodology provides competitive power. Besides, our empirical

power is strong even if the sample size is not sufficient (0.962 power when $n = 1500$). MMD test consumes least computation cost, but the empirical power was not satisfactory. KS test shows an average of 0 power, a possible reason makes the power so low is that f and g have similar mean (1.667 and 1.33, respectively), KS test also achieves low power in Table $A.13$, $A.14$, $A.17$, $A.19$, $A.21$, $A.29$, $A.30$, $A.31$, $A.32$, $A.33$, $A.34$, $A.35$, $A.36$, $A.37$, $A.43$ and $A.50$ when the means are similar. The Anderson-Darling test has similar empirical power to ours, but it consumes 1927.14 seconds when $n = 1500$, where our test uses 8.62 seconds, needless to say the computation cost when sample size increases. On the whole, our test can be the first choice to differentiate such samples.

Table 2.3. $f \sim$ t-distribution with $df = 2$ and $g \sim$ t-distribution with $df = 3$.

n	$\lambda_1(\tilde{A})$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.950 (7.72)	0.000 (3.44)	0.284 (2.22)	0.988 (2127.62)	0.046 (8.04)
2000	0.974 (14.40)	0.000 (3.98)	0.230 (2.95)	1.00 (3505.46)	0.042 (12.42)
2500	0.988 (19.09)	0.000 (4.52)	0.358 (3.52)	1.00 (8425.27)	0.044 (15.74)
3000	0.994 (23.05)	0.000 (5.21)	0.420 (3.86)	1.00 (17296.90)	0.052 (19.00)
3500	0.998 (27.54)	0.000 (5.63)	0.402 (4.62)	1.00 (23526.06)	0.044 (21.19)
4000	1.000 (32.29)	0.000 (5.99)	0.354 (5.05)	1.00 (29922.78)	0.058 (24.63)
4500	1.000 (36.05)	0.004 (6.32)	0.392 (5.29)	1.00 (35627.88)	0.044 (27.70)
5000	0.998 (38.82)	0.000 (7.95)	0.388 (6.93)	1.00 (43421.38)	0.056 (29.95)

Table 2.2 shows the empirical power when f and g are t-distribution with different parameters. Overall, all the empirical power are similar to the ones in table 2.2 except for MWU test, which achieves comparable power to KS test, a possible reason that results in small power is that both mean and variances between f and g are similar. MWU test also achieves low power in table $A.12$, $A.14$, $A.15$, $A.16$, $A.17$, $A.21$, $A.22$, $A.23$, $A.32$, $A.33$, $A.34$, $A.35$, $A.36$, $A.37$, $A.49$ and $A.50$ when two distributions have similar parameters. On the contrast, our test provides competitive power with low cost of computation.

In Figure 2.2, we introduced that t-distribution will become more centered if the degree of freedom enlarges. However, when both degree of freedom (df) are large, the distribution patterns between f and g can be resembled. Two samples are now harder to be distinguished apart. Thus, we also develop Table $A.34, A.35$ and $A.36$ to discuss the empirical power under such situation. As a result, our methodology achieves good empirical power and low computation cost in Table $A.34$.

Table 2.4. $f \sim$ Normal distribution (1,1) and $g \sim$ Normal (0.6,1).

n	$\lambda_1(\tilde{A})$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	Anderson-Darling (time elapsed(s))	MWU (time elapsed(s))
1500	0.954 (7.15)	1.000 (3.24)	0.342 (2.93)	0.974 (1883.25)	1.000 (8.29)
2000	0.992 (9.24)	1.000 (4.03)	0.362 (3.54)	0.988 (2793.73)	1.000 (12.35)
2500	0.998 (13.73)	1.000 (5.85)	0.412 (3.93)	1.000 (12693.26)	1.000 (14.83)
3000	1.000 (16.09)	1.000 (6.95)	0.426 (4.47)	1.000 (16732.95)	1.000 (18.50)
3500	1.000 (18.39)	1.000 (7.46)	0.518 (4.53)	1.000 (23902.52)	1.000 (22.63)
4000	1.000 (22.87)	1.000 (7.99)	0.542 (6.28)	1.000 (30931.64)	1.000 (25.89)
4500	1.000 (27.93)	1.000 (8.63)	0.584 (6.90)	1.000 (40293.09)	1.000 (28.63)
5000	1.000 (31.05)	1.000 (9.12)	0.590 (7.98)	1.000 (48392.75)	1.000 (30.82)

Table 2.4 shows the empirical power when f and g are Normal distribution with different parameters. Both KS test and MWU test suggest good power even the sample size is small (1500). The empirical power achieved from our methodology is also sufficient, which is 0.954 under same sample size. The computation cost is close to MWU test. Though MMD test still consumes the least computation cost, the largest empirical power computed is only 0.590, which is not satisfactory.

We also change the parameters in f and g, the empirical powers are shown in Table $A.22$, $A.23$, $A.24$, $A.26$ and $A.27$. Our two sample test has good power and low computation cost. According to Table $A.22$ and $A.23$, we discover that KS test achieves good empirical power when both samples are from Normal distribution, even if f and g have similar parameters. When both mean and variances are different between f and g, the MWU test can conclude better empirical power than KS test (see Table $A.26$ and $A.27$).

Table 2.5. $f \sim$ Cauchy-distribution (1,1.6) and $g \sim$ Cauchy-distribution (1.3,1.2).

n	$\lambda_1(\tilde{A})$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	Anderson-Darling (time elapsed(s))	MWU (time elapsed(s))
1500	1.000 (8.41)	0.488 (5.17)	0.298 (2.17)	0.994 (1912.79)	0.562 (9.01)
2000	1.000 (11.14)	0.644 (6.72)	0.284 (3.77)	0.996 (3724.72)	0.672 (12.54)
2500	1.000 (14.55)	0.752 (7.40)	0.320 (3.60)	0.998 (12923.96)	0.742 (14.99)
3000	1.000 (19.86)	0.850 (8.99)	0.384 (4.20)	0.998 (17527.12)	0.884 (18.33)
3500	1.000 (24.93)	0.866 (9.62)	0.420 (5.99)	0.998 (28421.23)	0.924 (20.82)
4000	1.000 (27.99)	0.910 (10.62)	0.446 (6.43)	1.00 (34882.72)	0.962 (29.90)
4500	1.000 (32.33)	0.942 (11.32)	0.472 (7.11)	1.00 (39723.19)	0.968 (31.19)
5000	1.000 (39.25)	0.982 (13.62)	0.488 (8.10)	1.00 (47231.42)	0.998 (35.62)

Table 2.5 shows the empirical power when f and g are Cauchy distribution with different parameters. In this case, our test achieves the most competitive empirical power. KS test and MWU test present good power when sample sizes are sufficient ($n \geq 3500$). MMD test still lacks strong empirical power under $\mathcal{H}_1$. The power from AD test is significant, but it still consumes the highest computational cost.

We change the parameters from both f and g to compute the empirical power again, which can be found in Table $A.14$, $A.15$, $A.16$, $A.17$, $A.18$, $A.19$, $A.20$ and $A.21$. Our methodology consistently proposes good empirical power and low computation cost. As discussed in Table 2.2, KS test possibly achieves low empirical power when f and g have similar mean. From Table $A.14$, $A.17$, $A.19$ and $A.21$, KS test suggests low power at an average of 0.

2.4.2. Samples from different distributions

In this part, we will discuss the empirical power and computation cost when f and g are under different distributions.

Table 2.6. $f \sim$ Normal-distribution (0,1) and $g \sim$ t-distribution $df = 2$.

n	$\lambda_1(\tilde{A})$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.978 (8.84)	0.564 (3.12)	0.352 (2.09)	0.952 (1997.25)	0.048 (9.31)
2000	1.000 (11.21)	0.620 (3.47)	0.362 (2.62)	1.00 (3811.38)	0.062 (13.47)
2500	1.000 (14.05)	0.666 (3.95)	0.404 (3.01)	1.00 (8523.75)	0.106 (16.52)
3000	1.000 (18.92)	0.648 (4.12)	0.490 (3.52)	1.00 (15392.23)	0.132 (18.01)
3500	1.000 (23.85)	0.712 (4.59)	0.520 (4.04)	1.00 (21333.12)	0.154 (21.32)
4000	1.000 (27.20)	0.800 (5.99)	0.534 (5.83)	1.00 (27392.54)	0.162 (24.53)
4500	1.000 (32.28)	0.832 (6.33)	0.558 (5.39)	1.00 (36496.37)	0.182 (28.35)
5000	1.000 (37.93)	0.878 (6.52)	0.560 (5.99)	1.00 (44238.90)	0.190 (27.99)

Table 2.6 suggests the empirical power when $f \sim$ Normal-distribution and $g \sim$ t-distribution, both distribution patterns are similar but g has fatter tails. In this case, the empirical powers calculated from MWU test are below 0.2, which are not satisfactory. KS test proposes good power when sample size is sufficient (In this case, $n > 4000$). MMD test doesn't achive strong empirical power. Our two sample test achieves the most competitive power, which is even better than the power computed from AD test. Besides, the computation cost is close to the MWU test, which is not costly.

Similar conclusion can be found in Table $A.50$, $A.51$, $A.52$ and $A.53$ when we change both distributions (f and g).

Table 2.7. $f \sim$ Chi-square distribution $df = 2$ and $g \sim$ F-distribution (2,4).

n	$\lambda_1(\tilde{A})$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
150	0.688 (0.41)	0.428 (1.84)	0.654 (346.31)	0.875 (292.43)	0.714 (0.72)
300	0.782 (0.80)	0.694 (2.29)	0.732 (1210.25)	0.928 (1062.89)	0.803 (1.21)
450	0.874 (0.96)	0.840 (2.07)	0.858 (2742.03)	0.945 (1306.07)	0.859 (1.36)
600	0.952 (1.25)	0.924 (3.92)	0.934 (4738.86)	1.000 (4296.68)	0.939 (1.99)
750	0.982 (2.48)	0.980 (3.63)	0.992 (7189.28)	1.000 (6274.58)	0.994 (2.21)
900	1.000 (2.86)	1.000 (3.50)	1.000 (10598.49)	1.000 (9528.81)	1.000 (2.89)

Table 2.7 presents the empirical power when $f \sim$ Chi-square and $g \sim$ F-distribution. In this case, MMD test uses the unbiased statistic (by Equation 1.5). Based on the outcome, MMD test now has stronger empirical power than previous ones. It has good power when n is sufficient (over 450). However, the computation cost drastically increases, which is now comparable to the AD test. For our test, we present good power with the least calculation time. Table $A.53$ also shows the empirical power when MMD uses unbiased statistic, the conclusion is similar.

As a conclusion, our two sample test achieves competitive power and low computation cost under $\mathcal{H}_1$. When distribution patterns between f and g are strongly resembled, our methodology still achieves satisfactory power. In addition, KS test and MWU test sometimes present 0 (or near 0) empirical power. However, our two sample test consistently achieves acceptable power. Overall, our test can present comparable power to AD test, and low computation cost that is similar to MWU test.

3. APPLICATION TO REAL DATA

Pancreatic cancer is one of the most lethal cancer type in U.S. As a common sense, male and female have different resistance to cancer. In other words, the probability distributions of surviving, after diagnosis of cancer, are different between male and female. This is also formally proved by Ilić and Ilic 2016. We use SEER (Surveillance, Epidemiology, and End Results) to collect pancreatic cancer data. Data includes patients who diagnosed to pancreatic cancer from year 2004 to 2015, and all the moralities occurred in stage 3 during the follow-up study (namely, all the observations are uncensored). It has 6659 female records and 7104 male records (13763 records in total). We focus on two variables, they are sex and survival months. The null hypothesis test with its alternative hypothesis test are:

$\mathcal{H}_0$: Male and female have same probability distribution of surviving after diagnosis of pancreatic cancer,

$\mathcal{H}_1$: Male and female have different probability distribution of surviving after diagnosis of pancreatic cancer.

Table 3.1. The p-value from the respective test under $\mathcal{H}_0$.

n	$\lambda_1(\tilde{A})$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
13763	0.000 (0.33)	0.047 (0.08)	0.642 (0.10)	0.075 (19.41)	0.150 (0.11)

From table 3.1, our two-sample test achieves the lowest (0.000) p-value under $\mathcal{H}_0$, which means we have very sufficient evidence to reject the null hypothesis under nominal level. Thus, we can conclude that male and female have different probability distribution of surviving after diagnosis of pancreatic cancer. Ilić and Ilic 2016 proposes same conclusion. Besides, the computational cost is also very low, which is similar to the cost from MWU test. Overall, our two sample test is the first choice for this real data.

4. CONCLUSION

Our two sample test is overall practical. For empirical type I error, our test shows an average of 0.04 in most distributions, no matter they are continuous or discrete. Besides, when the sample sizes are not sufficient, our methodology still achieves acceptable type I error. This is a good prerequisite for calculating empirical power. Under alternative hypothesis, our test consistently achieves good empirical power, which is comparable to the power calculated from AD test. When two distributions have resembled patterns, our test can still suggest good empirical power if the sample sizes are sufficient. Sometimes other tests can achieve zero power when two distributions have similar mean, but our empirical power is always acceptable. For the computation cost, our methodology consumes comparable cost to MWU test. To sum up, our two sample test has low computation cost, well controlled type I error and competitive empirical power, which can be a good choice to be applied to real data.

REFERENCES

Anderson, T. W., & Darling, D. (1952). Asymptotic theory of certain goodness of fit criteria based on stochastic processes. *Annals of Mathematical Statistics*, *23*, 193–212.

Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein generative adversarial networks. *Proceedings of Machine Learning Research*, *70*, 214–223.

Bickel, P. J., & Sarkar, P. (2016). Hypothesis testing for automated community detection in networks. *Journal of the Royal Statistical Society*, *78*, 253–273.

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

Hassani, H., & Silva, E. S. (2015). A kolmogorov-smirnov based test for comparing thepredictive accuracy of two sets of forecasts. *Econometrics*, *3*, 590–609.

Ilić, M., & Ilic, M. (2016). Epidemiology of pancreatic cancer. *World Journal of Gastroenterology*, *22*, 9694.

Kruskal, W. H. (1957). Historical notes on the wilcoxon unpaired two-sample test. *Journal of the American Statistical Association*, *52*, 356–360.

Lee, J. O., & Yin, J. (2014). A necessary and sufficient condition for edge universality of wigner matrices. *Duke Mathematical Journal*, *163*, 117–173.

Liu, Z., Xia, X., & Zhou, W. (2015). A test for equality of two distributions via jackknife empirical likelihood and characteristic functions. *Computational Statistics Data Analysis*, *92*, 97–114.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, *18*, 50–60.

Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society*, *231*, 289–337.

Pettitt, A. N. (1976). A Two-Sample Anderson-Darling Rank Statistic. *Biometrika*, *63*, 161–168.

Soshnikov, A. (1999). Universality at the edge of the spectrum in wigner random matrices. *Communications in Mathematical Physics*, *207*, 697–733.

Student. (1908). The probable error of a mean. *Biometrika*, *6*, 1–25.

Tracy, C. A., & Widom, H. (1993). Level-spacing distributions and the airy kernel. *Physics Letters*, *305*, 115–118.

Xu, M., Jog, V., & Loh, P.-L. (2017). Optimal rates for community estimation in the weighted stochastic block model. *The Annals of Statistics*, *48*, 183–204.

APPENDIX

Table A.1. Empirical type I error under Normal distribution.

n	(μ,σ)=(0,1)	(μ,σ)=(1,1)	(μ,σ)=(2,2)	(μ,σ)=(5,5)
150	0.021	0.016	0.012	0.013
550	0.030	0.036	0.027	0.034
950	0.029	0.038	0.047	0.042
1350	0.040	0.028	0.039	0.044
1750	0.028	0.035	0.038	0.046
2150	0.038	0.029	0.021	0.028
2550	0.041	0.038	0.035	0.032
2950	0.049	0.042	0.042	0.039
3350	0.037	0.041	0.047	0.041
3750	0.049	0.042	0.041	0.039

Table A.2. Empirical type I error under t-distribution.

n	$df = 5$	$df = 6$	$df = 7$	$df = 8$
150	0.034	0.022	0.026	0.018
550	0.082	0.050	0.034	0.024
950	0.100	0.049	0.046	0.018
1350	0.100	0.067	0.055	0.037
1750	0.101	0.070	0.032	0.051
2150	0.095	0.061	0.050	0.039
2550	0.117	0.051	0.041	0.024
2950	0.113	0.087	0.037	0.046
3350	0.114	0.059	0.021	0.030
3750	0.112	0.063	0.046	0.021

Table A.3. Empirical type I error under Gamma-distribution.

n	(α,β)=(2,0.1)	(α,β)=(2,0.2)	(α,β)=(2,0.3)	(α,β)=(2.5,0.3)
150	0.026	0.016	0.020	0.026
550	0.034	0.022	0.022	0.034
950	0.032	0.024	0.040	0.046
1350	0.026	0.040	0.032	0.032
1750	0.048	0.034	0.042	0.046
2150	0.040	0.042	0.046	0.048
2550	0.042	0.046	0.032	0.042
2950	0.038	0.042	0.050	0.048
3350	0.042	0.052	0.044	0.032
3750	0.046	0.044	0.042	0.048

Table A.4. Empirical type I error under Binomial-distribution.

n	(m,p)=(100 , 0.2)	(m,p)=(100 , 0.3)	(m,p)=(100,0.6)	(m,p)=(200,0.3)
150	0.020	0.028	0.016	0.020
550	0.030	0.028	0.026	0.030
950	0.040	0.032	0.038	0.040
1350	0.038	0.050	0.042	0.038
1750	0.042	0.064	0.030	0.042
2150	0.044	0.032	0.052	0.044
2550	0.048	0.030	0.046	0.048
2950	0.054	0.052	0.034	0.054
3350	0.050	0.054	0.022	0.034
3750	0.054	0.058	0.046	0.022

Table A.5. Empirical type I error under Weibull-distribution..

n	(α,β)=(3,1)	(α,β)=(3,1.2)	(α,β)=(2,1)	(α,β)=(3,2)
150	0.016	0.022	0.020	0.016
550	0.032	0.024	0.034	0.042
950	0.034	0.044	0.044	0.028
1350	0.026	0.032	0.030	0.036
1750	0.032	0.040	0.040	0.040
2150	0.038	0.054	0.040	0.048
2550	0.040	0.048	0.034	0.048
2950	0.040	0.042	0.036	0.052
3350	0.038	0.044	0.032	0.022
3750	0.032	0.040	0.038	0.028

Table A.6. Empirical type I error under Beta-distribution.

n	(α,β)=(3,3)	(α,β)=(3,4)	(α,β)=(3,5)	(α,β)=(4,3)
150	0.036	0.024	0.014	0.024
550	0.062	0.042	0.038	0.026
950	0.056	0.038	0.034	0.034
1350	0.040	0.034	0.020	0.038
1750	0.030	0.042	0.036	0.042
2150	0.046	0.040	0.042	0.058
2550	0.050	0.038	0.048	0.040
2950	0.044	0.042	0.052	0.032
3350	0.030	0.038	0.036	0.048
3750	0.036	0.034	0.048	0.056

Table A.7. Empirical type I error under Chi-square distribution.

n	$df = 5$	$df = 6$	$df = 7$	$df = 8$
150	0.020	0.026	0.022	0.016
550	0.030	0.036	0.022	0.034
950	0.038	0.026	0.042	0.032
1350	0.048	0.036	0.032	0.022
1750	0.048	0.048	0.026	0.034
2150	0.046	0.052	0.028	0.034
2550	0.024	0.032	0.044	0.046
2950	0.040	0.040	0.032	0.042
3350	0.034	0.032	0.026	0.020
3750	0.028	0.036	0.030	0.034

Table A.8. Empirical type I error under Exponential distribution.

n	β=1	β=1.5	β=2	β=2.5
150	0.048	0.060	0.052	0.068
550	0.094	0.098	0.072	0.092
950	0.076	0.098	0.104	0.068
1350	0.048	0.072	0.090	0.094
1750	0.104	0.110	0.080	0.072
2150	0.072	0.090	0.082	0.082
2550	0.084	0.096	0.062	0.096
2950	0.028	0.092	0.078	0.082
3350	0.090	0.102	0.092	0.086
3750	0.102	0.074	0.096	0.102

Table A.9. Empirical type I error under Poisson distribution.

n	λ=5	λ=6	λ=7	λ=8
150	0.026	0.024	0.012	0.018
550	0.018	0.038	0.020	0.032
950	0.034	0.034	0.030	0.040
1350	0.030	0.032	0.052	0.046
1750	0.042	0.030	0.042	0.044
2150	0.032	0.040	0.042	0.042
2550	0.036	0.034	0.040	0.042
2950	0.038	0.050	0.042	0.038
3350	0.028	0.038	0.038	0.042
3750	0.040	0.044	0.042	0.042

Table A.10. Empirical type I error under Mixed normal distribution.

n	(μ_1,μ_2)=(0.2 ,0.4) $(\mathrm{p}_1, p_2) = (0.4, 0.6)$ (σ_1,σ_2)=(1,0.9)	(μ_1,μ_2)=(0.3,0.6) $(\mathrm{p}_1, p_2) = (0.4, 0.6)$ (σ_1,σ_2)=(1,0.9)	(μ_1,μ_2)=(0.2,0.4) $(\mathrm{p}_1, p_2) = (0.5, 0.5)$ (σ_1,σ_2)=(1,0.9)	(μ_1,μ_2)=(0.2,0.4) $(\mathrm{p}_1, p_2) = (0.4, 0.6)$ (σ_1,σ_2)=(2,2.2)
150	0.012	0.028	0.022	0.018
550	0.028	0.034	0.026	0.042
950	0.020	0.038	0.046	0.032
1350	0.032	0.024	0.032	0.042
1750	0.040	0.036	0.040	0.042
2150	0.038	0.024	0.020	0.040
2550	0.032	0.030	0.036	0.038
2950	0.034	0.044	0.044	0.044
3350	0.038	0.044	0.052	0.042
3750	0.042	0.046	0.048	0.038

Table A.11. Empirical power under $f \sim$ F-distribution (df_1,df_2)=(2,5) and $g \sim$ F-distribution (df_1,df_2) =(2,8).

n	$\lambda_1(\tilde{A})$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.962 (8.62)	0.020 (3.03)	0.292 (2.17)	0.962 (1927.14)	0.346 (9.01)
2000	0.970 (11.02)	0.000 (3.45)	0.230 (2.80)	1.000 (3749.28)	0.416 (11.57)
2500	0.989 (15.50)	0.020 (3.99)	0.264 (3.08)	1.000 (8202.25)	0.500 (14.47)
3000	1.000 (19.38)	0.000 (4.04)	0.306 (3.68)	1.000 (14092.02)	0.612 (17.74)
3500	0.990 (25.17)	0.000 (4.39)	0.334 (4.24)	1.000 (20175.98)	0.700 (20.03)
4000	1.000 (29.29)	0.000 (5.9)	0.358 (5.96)	1.000 (26602.84)	0.714 (22.80)
4500	1.000 (35.94)	0.000 (6.2)	0.342 (5.20)	1.000 (34496.27)	0.764 (25.49)
5000	1.000 (40.92)	0.020 (5.95)	0.370 (5.72)	1.000 (41014.28)	0.872 (26.75)

Table A.12. Empirical power under $f \sim$ Normal distribution (μ=0,σ=1) and $g \sim$ t-distribution ($df = 2$).

n	$\tilde{\lambda_1(A)}$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.979 (8.84)	0.564 (3.12)	0.352 (2.09)	0.9532 (1997.25)	0.048 (9.31)
2000	1.000 (11.21)	0.620 (3.47)	0.362 (2.62)	1.00 (3811.38)	0.062 (13.47)
2500	1.000 (14.05)	0.666 (3.95)	0.404 (3.01)	1.00 (8523.75)	0.105 (16.52)
3000	1.000 (18.92)	0.648 (4.12)	0.490 (3.52)	1.00 (15392.23)	0.131 (18.01)
3500	1.000 (23.85)	0.712 (4.59)	0.520 (4.04)	1.00 (21333.12)	0.152 (21.32)
4000	1.000 (27.20)	0.800 (5.99)	0.534 (5.83)	1.00 (27392.54)	0.162 (24.53)
4500	1.000 (32.28)	0.832 (6.33)	0.558 (5.39)	1.00 (36496.37)	0.183 (28.35)
5000	1.000 (37.93)	0.878 (6.52)	0.560 (5.99)	1.00 (44238.90)	0.191 (27.99)

Table A.13. Empirical power under $f \sim$ t-distribution (df=2) and $g \sim$ t-distribution (df=3).

n	$\tilde{\lambda_1(A)}$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.950 (7.72)	0.000 (3.44)	0.284 (2.22)	0.988 (2127.62)	0.046 (8.04)
2000	0.974 (14.40)	0.000 (3.98)	0.230 (2.95)	1.00 (3505.46)	0.042 (12.42)
2500	0.988 (19.09)	0.000 (4.52)	0.358 (3.52)	1.00 (8425.27)	0.044 (15.74)
3000	0.994 (23.05)	0.000 (5.21)	0.420 (3.86)	1.00 (17296.90)	0.052 (19.00)
3500	0.998 (27.54)	0.000 (5.63)	0.402 (4.62)	1.00 (23526.06)	0.044 (21.19)
4000	1.000 (32.29)	0.000 (5.99)	0.354 (5.05)	1.00 (29922.78)	0.058 (24.63)
4500	1.000 (36.05)	0.004 (6.32)	0.392 (5.29)	1.00 (35627.88)	0.044 (27.70)
5000	0.998 (38.82)	0.000 (7.95)	0.388 (6.93)	1.00 (43421.38)	0.056 (29.95)

Table A.14. Empirical power under $f \sim$ Cauchy distribution (α,β)=(0,1) and $g \sim$ Cauchy distribution (α,β)=(0,1.2).

n	$\tilde{\lambda_1(A)}$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.998 (9.61)	0.000 (4.02)	0.294 (3.16)	0.962 (2318.52)	0.058 (10.00)
2000	0.998 (18.39)	0.000 (3.82)	0.358 (3.17)	0.998 (3409.36)	0.036 (11.94)
2500	1.000 (20.80)	0.000 (4.28)	0.346 (3.38)	1.00 (8824.23)	0.058 (17.97)
3000	1.000 (31.61)	0.002 (4.56)	0.368 (3.71)	1.00 (19136.24)	0.056 (20.20)
3500	1.000 (36.34)	0.000 (5.01)	0.374 (4.41)	1.00 (25212.25)	0.062 (23.02)
4000	1.000 (47.51)	0.000 (5.52)	0.402 (6.01)	1.00 (33126.26)	0.054 (25.79)
4500	1.000 (53.24)	0.000 (6.15)	0.442 (6.93)	1.00 (37217.55)	0.034 (28.42)
5000	1.000 (54.71)	0.000 (6.75)	0.488 (6.52)	1.00 (46231.62)	0.050 (29.21)

Table A.15. Empirical power under $f \sim$ Cauchy distribution (α,β)=(0,1) and $g \sim$ Cauchy distribution (α,β)=(0,1.6).

n	$\lambda_1(\tilde{A})$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.998 (9.57)	0.230 (3,98)	0.320 (3.12)	0.988 (2215.72)	0.032 (9.96)
2000	1.000 (18.32)	0.298 (3.65)	0.394 (3.10)	0.994 (3268.48)	0.052 (11.87)
2500	1.000 (21.15)	0.416 (4.64)	0.358 (3.73)	1.00 (9214.24)	0.046 (15.12)
3000	1.000 (32.11)	0.510 (4.77)	0.346 (4.41)	1.00 (18772.48)	0.056 (18.47)
3500	1.000 (36.20)	0.626 (4.42)	0.356 (4.87)	1.00 (26231.72)	0.070 (20.06)
4000	0.998 (47.65)	0.732 (6.16)	0.422 (5.01)	1.00 (34268.25)	0.038 (23.16)
4500	1.000 (53.26)	0.764 (6.52)	0.398 (5.72)	1.00 (39276.88)	0.046 (25.81)
5000	1.000 (54.12)	0.818 (6.99)	0.386 (6.15)	1.00 (45282.90)	0.048 (26.95)

Table A.16. Empirical power under $f \sim$ Cauchy distribution (α,β)=(0,1) and $g \sim$ Cauchy distribution (α,β)=(0,1.8).

n	$\lambda_1(\tilde{A})$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	1.000 (9.59)	0.774 (4.00)	0.302 (3.14)	0.994 (2325.10)	0.036 (9.98)
2000	1.000 (18.12)	0.948 (3.55)	0.334 (2.90)	0.996 (3125.22)	0.056 (11.23)
2500	1.000 (20.71)	0.980 (4.20)	0.316 (3.29)	1.00 (9904.15)	0.046 (14.72)
3000	1.000 (31.77)	0.992 (4.43)	0.380 (4.07)	1.00 (17882.77)	0.038 (19.24)
3500	1.000 (35.90)	1.000 (4.86)	0.362 (4.71)	1.00 (27261.18)	0.056 (22.15)
4000	1.000 (49.52)	1.000 (6.50)	0.360 (6.56)	1.00 (35215.26)	0.062 (26.15)
4500	1.000 (57.25)	1.000 (7.19)	0.342 (6.17)	1.00 (39929.27)	0.050 (28.82)
5000	1.000 (55.11)	1.000 (6.76)	0.392 (6.53)	1.00 (48231.52)	0.062 (32.50)

Table A.17. Empirical power under $f \sim$ Cauchy distribution (α,β)=(1,1.2) and $g \sim$ Cauchy distribution (α,β)=(1,1.6).

n	$\lambda_1(\tilde{A})$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.996 (8.68)	0.000 (3.09)	0.3314 (2.23)	0.992 (1993.14)	0.050 (9.07)
2000	1.000 (18.06)	0.000 (3.49)	0.312 (2.84)	0.998 (3765.14)	0.054 (11.61)
2500	1.000 (21.40)	0.002 (4.89)	0.314 (3.98)	1.00 (9924.72)	0.042 (15.37)
3000	0.998 (32.37)	0.002 (5.03)	0.318 (4.67)	1.00 (17442.16)	0.062 (18.73)
3500	1.000 (36.67)	0.000 (4.90)	0.318 (4.74)	1.00 (26252.62)	0.046 (20.53)
4000	1.000 (47.99)	0.004 (6.60)	0.320 (6.66)	1.00 (36234.14)	0.058 (23.50)
4500	1.000 (52.59)	0.000 (6.22)	0.360 (6.52)	1.00 (39727.96)	0.046 (25.51)
5000	1.000 (59.42)	0.000 (6.27)	0.348 (7.12)	1.00 (46276.72)	0.062 (31.12)

Table A.18. Empirical power under $f \sim$ Cauchy distribution (α,β)=(0.9,1) and $g \sim$ Cauchy distribution (α,β)=(1,1.4).

n	$\lambda_1(\tilde{A})$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	1.000 (8.74)	0.080 (3.15)	0.3314 (2.53)	0.994 (1723.72)	0.050 (9.52)
2000	1.000 (18.30)	0.110 (3.73)	0.312 (3.44)	0.994 (36762.62)	0.104 (11.62)
2500	1.000 (21.41)	0.128 (5.02)	0.304 (4.00)	1.00 (1004.26)	0.196 (15.80)
3000	1.000 (36.59)	0.162 (5.94)	0.418 (4.55)	1.00 (17252.66)	0.242 (18.62)
3500	1.000 (39.80)	0.192 (6.15)	0.402 (4.94)	1.00 (27252.02)	0.346 (20.47)
4000	0.998 (46.65)	0.210 (6.68)	0.422 (5.72)	1.00 (35272.67)	0.458 (23.48)
4500	1.000 (49.92)	0.230 (7.03)	0.388 (6.30)	1.00 (39932.30)	0.568 (25.28)
5000	1.000 (54.72)	0.332 (7.52)	0.422 (7.55)	1.00 (46288.72)	0.698 (31.21)

Table A.19. Empirical power under$f \sim$ Cauchy distribution (α,β)=(1,1) and $g \sim$ Cauchy distribution (α,β)=(1,1.2).

n	$\lambda_1(\tilde{A})$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	1.000 (9.03)	0.000 (3.44)	0.292 (2.58)	0.990 (1942.29)	0.896 (9.12)
2000	0.998 (18.66)	0.000 (4.09)	0.328 (3.44)	0.992 (3662.12)	0.924 (11.03)
2500	0.998 (20.61)	0.002 (4.10)	0.420 (3.92)	0.998 (12004.72)	0.946 (14.62)
3000	1.000 (32.33)	0.000 (5.12)	0.354 (4.72)	1.00 (17722.39)	0.952 (18.52)
3500	1.000 (36.78)	0.000 (6.29)	0.464 (5.02)	1.00 (28232.28)	0.964 (20.32)
4000	0.998 (47.68)	0.000 (6.38)	0.462 (5.52)	1.00 (35882.29)	0.984 (27.32)
4500	0.998 (53.12)	0.000 (6.62)	0.352 (6.28)	1.00 (37823.88)	0.992 (28.25)
5000	1.000 (54.29)	0.000 (7.10)	0.390 (7.72)	1.00 (46214.79)	0.988 (33.42)

Table A.20. Empirical power under $f \sim$ Cauchy distribution (α,β)=(1.3,1.2) and $g \sim$ Cauchy distribution (α,β)=(1,1.6).

n	$\lambda_1(\tilde{A})$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	1.000 (8.41)	0.488 (5.17)	0.298 (2.17)	0.994 (1912.79)	0.562 (9.01)
2000	1.000 (11.14)	0.644 (6.72)	0.284 (3.77)	0.996 (3724.72)	0.672 (12.54)
2500	1.000 (14.55)	0.752 (7.40)	0.320 (3.60)	0.998 (12923.96)	0.742 (14.99)
3000	1.000 (19.86)	0.850 (8.99)	0.384 (4.20)	0.998 (17527.12)	0.884 (18.33)
3500	1.000 (24.93)	0.866 (9.62)	0.420 (5.99)	0.998 (28421.23)	0.924 (20.82)
4000	1.000 (27.92)	0.910 (10.62)	0.446 (6.43)	1.00 (34882.72)	0.962 (29.90)
4500	1.000 (32.33)	0.942 (11.32)	0.472 (7.11)	1.00 (39723.19)	0.968 (31.19)
5000	1.000 (39.25)	0.982 (13.62)	0.488 (8.10)	1.00 (47231.42)	0.998 (35.62)

Table A.21. Empirical power under f ∼ Cauchy distribution (α,β)=(1.2,2.4) and g ∼ Cauchy distribution (α,β)=(1.3,2).

n	$\lambda_1(\tilde{A})$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.998 (9.3)	0.002 (3.71)	0.284 (2.85)	0.948 (2123.19)	0.056 (9.69)
2000	1.000 (18.81)	0.004 (4.24)	0.316 (3.59)	0.982 (3562.03)	0.094 (12.36)
2500	1.000 (22.76)	0.000 (4.25)	0.282 (3.34)	0.998 (15823.12)	0.102 (14.73)
3000	1.000 (26.95)	0.000 (4.63)	0.298 (4.27)	0.998 (19239.05)	0.124 (18.33)
3500	1.000 (32.13)	0.000 (4.70)	0.302 (5.42)	1.000 (27293.85)	0.188 (20.34)
4000	1.000 (43.52)	0.020 (6.00)	0.348 (6.06)	1.000 (35293.02)	0.186 (24.13)
4500	1.000 (52.53)	0.000 (6.62)	0.368 (5.62)	1.000 (41233.00)	0.212 (25.86)
5000	1.000 (55.08)	0.000 (6.19)	0.326 (6.95)	1.000 (48592.11)	0.224 (32.09)

Table A.22. Empirical power under f ∼ Normal distribution (μ,σ)=(0,1) and g ∼ Normal distribution (μ,σ)=(0,1.2).

n	$\lambda_1(\tilde{A})$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.036 (9.23)	0.906 (3.93)	0.264 (2.29)	0.948 (2323.85)	0.046 (9.37)
2000	0.046 (18.96)	0.984 (4.78)	0.294 (3.49)	0.982 (3425.63)	0.046 (13.33)
2500	0.034 (20.29)	0.996 (5.09)	0.226 (3.90)	0.998 (16632.93)	0.050 (16.72)
3000	0.032 (32.13)	1.000 (5.53)	0.312 (4.42)	1.000 (19739.09)	0.034 (18.63)
3500	0.028 (36.92)	1.000 (5.92)	0.394 (5.35)	1.000 (26293.83)	0.054 (21.63)
4000	0.026 (46.23)	1.000 (6.13)	0.300 (5.91)	1.000 (39503.92)	0.052 (24.35)
4500	0.034 (49.91)	1.000 (6.94)	0.402 (6.12)	1.000 (40103.38)	0.060 (27.88)
5000	0.034 (53.42)	1.000 (7.10)	0.416 (6.83)	1.000 (47932.06)	0.048 (30.14)

Table A.23. Empirical power under f ∼ Normal distribution (μ,σ)=(0,1) and g ∼ Normal distribution (μ,σ) =(0,2).

n	$\lambda_1(\tilde{A})$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.032 (9.83)	1.000 (4.05)	0.268 (3.07)	0.962 (2105.63)	0.060 (9.91)
2000	0.022 (15.15)	1.000 (4.89)	0.312 (3.36)	0.988 (3238.42)	0.068 (12.13)
2500	0.036 (18.06)	1.000 (5.23)	0.292 (3.53)	0.994 (13632.93)	0.066 (14.86)
3000	0.048 (24.95)	1.000 (5.96)	0.320 (4.37)	1.000 (17930.03)	0.056 (17.16)
3500	0.042 (29.94)	1.000 (6.26)	0.324 (4.32)	1.000 (23032.68)	0.080 (22.75)
4000	0.060 (35.10)	1.000 (6.85)	0.326 (6.84)	1.000 (36302.73)	0.066 (28.45)
4500	0.028 (39.32)	1.000 (7.63)	0.452 (5.29)	1.000 (42103.98)	0.078 (31.32)
5000	0.044 (43.93)	1.000 (8.63)	0.410 (6.53)	1.000 (48021.69)	0.080 (35.98)

Table A.24. Empirical power under $f \sim$ Normal distribution (μ,σ)=(1,1) and $g \sim$ Normal distribution (μ,σ) =(0,2).

n	$\tilde{\lambda_1(A)}$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	1.000 (8.18)	1.000 (3.79)	0.364 (2.91)	0.924 (2503.85	1.000 (9.68)
2000	1.000 (12.68)	1.000 (4.26)	0.342 (3.64)	0.958 (3793.73)	1.000 (15.22)
2500	1.000 (16.27)	1.000 (5.93)	0.494 (3.57)	0.998 (14637.09)	1.000 (18.26)
3000	1.000 (19.76)	1.000 (5.03)	0.326 (4.15)	1.000 (19832.76)	1.000 (21.48)
3500	1.000 (23.53)	1.000 (6.64)	0.524 (4.73)	1.000 (24932.72)	1.000 (26.75)
4000	1.000 (29.37)	1.000 (7.35)	0.328 (5.73)	1.000 (33932.04)	1.000 (29.04)
4500	1.000 (34.75)	1.000 (7.90)	0.450 (5.92)	1.000 (41032.85)	1.000 (36.73)
5000	1.000 (39.10)	1.000 (8.53)	0.552 (6.07)	1.000 (47039.95)	1.000 (39.08)

Table A.25. Empirical power under $f \sim$ Normal distribution (μ,σ)=(1,1) and $g \sim$ Normal distribution (μ,σ)=(0.6,1).

n	$\tilde{\lambda_1(A)}$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.954 (7.15)	1.000 (3.24)	0.342 (2.93)	0.974 (1883.25	1.000 (8.29)
2000	0.992 (9.24)	1.000 (4.03)	0.362 (3.54)	0.988 (2793.73)	1.000 (12.35)
2500	0.998 (13.73)	1.000 (5.85)	0.412 (3.93)	1.000 (12693.26)	1.000 (14.83)
3000	1.000 (16.09)	1.000 (6.95)	0.426 (4.47)	1.000 (16732.95)	1.000 (18.50)
3500	1.000 (18.39)	1.000 (7.46)	0.518 (4.53)	1.000 (23902.52)	1.000 (22.63)
4000	1.000 (22.87)	1.000 (7.99)	0.542 (6.28)	1.000 (30931.64)	1.000 (25.89)
4500	1.000 (27.93)	1.000 (8.63)	0.584 (6.90)	1.000 (40293.09)	1.000 (28.63)
5000	1.000 (31.05)	1.000 (9.12)	0.590 (7.98)	1.000 (48392.75)	1.000 (30.82)

Table A.26. Empirical power under $f \sim$ Normal distribution (μ,σ)=(0.4,1) and $g \sim$ Normal distribution (μ,σ)=(0.6,1.2).

n	$\tilde{\lambda_1(A)}$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.102 (7.10)	0.358 (2.09)	0.294 (1.58)	0.988 (2123.77)	1.000 (7.16)
2000	0.144 (9.83)	0.456 (2.88)	0.330 (3.09)	0.998 (2930.96)	1.000 (9.14)
2500	0.178 (15.72)	0.506 (4.16)	0.302 (5.73)	1.000 (15093.58)	1.000 (13.37)
3000	0.230 (18.32)	0.496 (6.23)	0.336 (6.73)	1.000 (18023.19)	1.000 (19.39)
3500	0.310 (21.87)	0.604 (7.86)	0.384 (8.26)	1.000 (24038.28)	1.000 (24.57)
4000	0.358 (25.83)	0.732 (8.87)	0.410 (10.53)	1.000 (369503.60)	1.000 (28.04)
4500	0.362 (29.20)	0.778 (10.15)	0.464 (13.76)	1.000 (42183.72)	1.000 (31.84)
5000	0.428 (35.75)	0.818 (12.76)	0.388 (16.73)	1.000 (49283.29)	1.000 (38.90)

Table A.27. Empirical power under $f \sim$ Normal distribution (μ,σ)=(2,2) and $g \sim$ Normal distribution (μ,σ)=(1.7,2.5).

n	$\tilde{\lambda_1}(A)$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.034 (7.63)	0.292 (4.02)	0.192 (3.16)	0.908 (1759.29)	0.802 (10.00)
2000	0.106 (10.52)	0.318 (3.85)	0.204 (4.64)	0.942 (2365.83)	0.878 (12.79)
2500	0.194 (14.62)	0.404 (4.40)	0.224 (5.08)	0.988 (9426.73)	0.914 (17.09)
3000	0.284 (18.21)	0.622 (4.65)	0.280 (6.98)	1.000 (15143.82)	0.960 (23.81)
3500	0.306 (22.09)	0.718 (6.99)	0.288 (9.12)	1.000 (21538.75)	1.000 (26.08)
4000	0.384 (26.73)	0.822 (8.83)	0.268 (12.08)	1.000 (36930.02)	1.000 (29.82)
4500	0.404 (28.19)	0.984 (11.95)	0.278 (15.86)	1.000 (41069.85)	1.000 (34.26)
5000	0.464 (33.90)	0.998 (13.69)	0.302 (19.04)	1.000 (47286.17)	1.000 (39.92)

Table A.28. Empirical power under $f \sim$ F-distribution (df_1, df_2)=(2,5) and $g \sim$ F-distribution (df_1, df_2) =(2,7).

n	$\tilde{\lambda_1}(A)$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.120 (9.36)	0.020 (4.02)	0.318 (2.09)	0.920 (1492.90)	0.108 (9.89)
2000	0.160 (12.28)	0.080 (3.85)	0.320 (3.73)	0.962 (4163.07)	0.212 (13.25)
2500	0.204 (12.82)	0.102 (4.40)	0.356 (5.63)	0.998 (9273.10)	0.268 (17.24)
3000	0.268 (14.09)	0.122 (4.65)	0.392 (7.26)	1.000 (13723.83)	0.342 (21.73)
3500	0.342 (19.79)	0.116 (6.99)	0.412 (9.69)	1.000 (22057.04)	0.406 (24.03)
4000	0.424 (23.08)	0.068 (8.83)	0.348 (13.63)	1.000 (37038.19)	0.428 (27.27)
4500	0.498 (26.73)	0.146 (11.95)	0.418 (16.73)	1.000 (42036.16)	0.486 (31.77)
5000	0.544 (31.75)	0.182 (13.69)	0.522 (21.63)	1.000 (48927.10)	0.502 (37.05)

Table A.29. Empirical power under $f \sim$ F-distribution(df_1, df_2)=(2,5) and $g \sim$ F-distribution(df_1, df_2) =(2,10).

n	$\tilde{\lambda_1}(A)$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.960 (7.85)	0.000 (2.79)	0.344 (3.15)	0.974 (1518.94)	0.444 (9.59)
2000	0.988 (9.58)	0.002 (3.57)	0.336 (3.89)	0.998 (5268.27)	0.564 (13.25)
2500	0.998 (12.28)	0.004 (5.85)	0.346 (5.28)	1.000 (14279.20)	0.648 (16.95)
3000	0.998 (17.27)	0.000 (7.58)	0.374 (7.69)	1.000 (19269.75)	0.734 (18.68)
3500	1.000 (19.85)	0.002 (11.87)	0.388 (9.28)	1.000 (26267.54)	0.814 (21.26)
4000	0.998 (23.69)	0.008 (15.84)	0.428 (12.58)	1.000 (36932.10)	0.850 (26.82)
4500	1.000 (26.27)	0.000 (18.62)	0.404 (16.46)	1.000 (39529.73)	0.900 (28.95)
5000	0.998 (36.83)	0.002 (21.87)	0.458 (19.94)	1.000 (46279.83)	0.918 (31.64)

Table A.30. Empirical power under $f \sim$ F-distribution(df_1, df_2)=(2,5) and $g \sim$ F-distribution(df_1, df_2) =(2,11).

n	$\tilde{\lambda_1}(A)$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.964 (9.36)	0.002 (2.66)	0.302 (3.52)	0.958 (1537.78)	0.544 (9.59)
2000	0.976 (10.53)	0.004 (4.13)	0.332 (4.37)	0.972 (4875.38)	0.642 (12.73)
2500	0.998 (16.79)	0.002 (5.69)	0.374 (4.97)	0.998 (13569.39)	0.776 (15.65)
3000	0.998 (18.80)	0.002 (9.27)	0.392 (6.69)	1.000 (19172.58)	0.798 (18.95)
3500	1.000 (21.63)	0.006 (13.69)	0.348 (8.63)	1.000 (23628.53)	0.828 (22.83)
4000	1.000 (25.43)	0.002 (16.49)	0.402 (11.73)	1.000 (31728.74)	0.884 (25.29)
4500	1.000 (25.82)	0.004 (19.47)	0.454 (15.83)	1.000 (37327.00)	0.924 (28.16)
5000	1.000 (31.19)	0.002 (21.70)	0.428 (18.26)	1.000 (43742.62)	0.982 (32.59)

Table A.31. Empirical power under $f \sim$ F-distribution(df_1, df_2)=(2,15) and $g \sim$ F-distribution(df_1, df_2) =(2,16).

n	$\tilde{\lambda_1}(A)$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.044 (8.99)	0.000 (2.66)	0.284 (2.54)	0.982 (1327.85)	0.544 (9.59)
2000	0.044 (15.25)	0.000 (4.13)	0.356 (3.18)	0.992 (4952.23)	0.642 (12.73)
2500	0.040 (17.83)	0.000 (5.69)	0.326 (4.09)	1.000 (16744.14)	0.776 (15.65)
3000	0.102 (25.73)	0.000 (9.27)	0.338 (6.25)	1.000 (19843.17)	0.798 (18.95)
3500	0.144 (28.21)	0.000 (13.69)	0.420 (8.23)	1.000 (24725.19)	0.828 (22.83)
4000	0.168 (31.84)	0.000 (16.49)	0.340 (12.84)	1.000 (29526.28)	0.884 (25.29)
4500	0.182 (36.63)	0.002 (19.47)	0.400 (17.32)	1.000 (32532.05)	0.924 (28.16)
5000	0.194 (38.27)	0.000 (21.70)	0.384 (19.73)	1.000 (41562.29)	0.982 (32.59)

Table A.32. Empirical power under $f \sim$ F-distribution(df_1, df_2)=(2,25) and $g \sim$ F-distribution(df_1, df_2) =(2,20).

n	$\tilde{\lambda_1}(A)$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.334 (9.58)	0.000 (3.73)	0.384 (3.85)	0.982 (1575.83)	0.050 (7.27)
2000	0.384 (12.62)	0.000 (4.84)	0.392 (5.95)	0.992 (4673.18)	0.076 (9.98)
2500	0.442 (15.84)	0.000 (5.27)	0.402 (8.21)	1.000 (17272.62)	0.072 (13.57)
3000	0.562 (19.63)	0.002 (9.84)	0.400 (12.53)	1.000 (21426.94)	0.062 (18.21)
3500	0.588 (24.29)	0.000 (15.62)	0.386 (14.72)	1.000 (25265.69)	0.066 (24.21)
4000	0.662 (29.73)	0.000 (19.21)	0.392 (18.21)	1.000 (29803.18)	0.052 (28.29)
4500	0.722 (32.79)	0.000 (23.62)	0.428 (20.73)	1.000 (33425.68)	0.066 (29.93)
5000	0.798 (35.79)	0.000 (25.93)	0.524 (22.73)	1.000 (44800.82)	0.072 (38.83)

Table A.33. Empirical power under f ∼ F-distribution(df_1, df_2)=(4,5) and g ∼ F-distribution (df_1, df_2) =(4,7).

n	$\tilde{\lambda_1}(A)$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.974 (9.15)	0.002 (2.95)	0.356 (3.11)	0.990 (1473.90)	0.050 (7.27)
2000	0.990 (11.83)	0.000 (2.88)	0.314 (4.92)	0.998 (4559.98)	0.076 (9.98)
2500	0.998 (13.87)	0.000 (3.79)	0.394 (6.53)	1.000 (18238.14)	0.072 (13.57)
3000	1.000 (16.62)	0.000 (4.37)	0.412 (9.25)	1.000 (23255.28)	0.062 (18.21)
3500	1.000 (19.42)	0.000 (6.26)	0.408 (11.26)	1.000 (28267.29)	0.066 (24.21)
4000	1.000 (24.91)	0.000 (8.82)	0.424 (14.74)	1.000 (31426.48)	0.052 (28.29)
4500	1.000 (28.10)	0.000 (14.62)	0.486 (19.39)	1.000 (35245.09)	0.066 (29.93)
5000	0.998 (32.18)	0.000 (19.82)	0.424 (21.03)	1.000 (43425.52)	0.072 (38.83)

Table A.34. Empirical power under f ∼ t-distribution (df=2) and g ∼ t-distribution (df=3).

n	$\tilde{\lambda_1}(A)$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.950 (8.48)	0.000 (3.66)	0.284 (3.29)	0.918 (1313.05)	0.046 (9.93)
2000	0.974 (10.31)	0.000 (4.21)	0.302 (4.44)	0.948 (4526.26)	0.042 (12.20)
2500	0.978 (14.39)	0.000 (4.63)	0.358 (5.10)	0.998 (16326.47)	0.052 (14.49)
3000	0.992 (17.70)	0.002 (5.73)	0.420 (6.11)	1.000 (24521.87)	0.044 (19.84)
3500	1.000 (20.67)	0.000 (6.25)	0.354 (8.84)	1.000 (29256.17)	0.058 (23.62)
4000	1.000 (23.59)	0.004 (7.99)	0.392 (11.57)	1.000 (34257.19)	0.044 (29.42)
4500	1.000 (26.53)	0.000 (9.14)	0.388 (15.73)	1.000 (39365.94)	0.042 (31.69)
5000	1.000 (35.84)	0.000 (13.53)	0.400 (18.09)	1.000 (44655.29)	0.084 (35.27)

Table A.35. Empirical power under f ∼ t-distribution (df=4) and g ∼ t-distribution (df=7).

n	$\tilde{\lambda_1}(A)$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.044 (9.58)	0.002 (4.00)	0.006 (3.34)	0.424 (1863.15)	0.064 (8.25)
2000	0.044 (11.93)	0.000 (3.78)	0.018 (4.19)	0.468 (4723.58)	0.062 (11.27)
2500	0.040 (13.27)	0.000 (4.58)	0.010 (5.75)	0.536 (15284.92)	0.036 (12.79)
3000	0.102 (16.29)	0.002 (5.15)	0.022 (6.63)	0.588 (23251.06)	0.038 (14.79)
3500	0.144 (19.94)	0.000 (7.09)	0.024 (7.88)	0.602 (28245.43)	0.048 (17.80)
4000	0.164 (24.12)	0.002 (8.58)	0.030 (9.74)	0.638 (33214.62)	0.062 (20.69)
4500	0.230 (27.03)	0.000 (10.05)	0.030 (13.66)	0.652 (38545.10)	0.068 (23.42)
5000	0.284 (31.82)	0.000 (14.48)	0.032 (17.59)	0.674 (43525.89)	0.050 (27.70)

Table A.36. Empirical power under $f \sim$ t-distribution (df=10) and $g \sim$ t-distribution (df=15).

n	$\tilde{\lambda_1}(A)$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.092 (3.35)	0.002 (3.27)	0.050 (2.25)	0.606 (2124.53)	0.042 (7.62)
2000	0.104 (3.95)	0.006 (3.73)	0.064 (3.57)	0.624 (5252.09)	0.046 (12.06)
2500	0.142 (12.85)	0.000 (4.93)	0.062 (4.68)	0.658 (16724.84)	0.044 (14.27)
3000	0.168 (14.59)	0.000 (5.58)	0.072 (5.05)	0.682 (25253.57)	0.058 (16.83)
3500	0.206 (16.09)	0.000 (6.72)	0.086 (6.73)	0.706 (29428.62)	0.054 (19.22)
4000	0.238 (20.46)	0.000 (8.29)	0.092 (8.83)	0.710 (34528.79)	0.066 (21.65)
4500	0.268 (24.62)	0.000 (9.84)	0.098 (11.68)	0.728 (39734.58)	0.070 (25.27)
5000	0.322 (28.94)	0.000 (12.63)	0.104 (15.75)	0.750 (45275.79)	0.082 (29.85)

Table A.37. Empirical power under $f \sim$ t-distribution (df=18) and $g \sim$ t-distribution (df=22).

n	$\tilde{\lambda_1}(A)$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.382 (3.35)	0.000 (3.45)	0.108 (2.58)	0.874 (1973.63)	0.064 (9.43)
2000	0.354 (3.95)	0.004 (3.94)	0.114 (3.52)	0.942 (4526.52)	0.040 (11.67)
2500	0.366 (12.85)	0.002 (4.25)	0.120 (4.55)	0.992 (13562.58)	0.038 (15.36)
3000	0.404 (14.59)	0.000 (6.52)	0.136 (6.28)	0.998 (24527.83)	0.062 (18.27)
3500	0.358 (16.09)	0.002 (7.43)	0.152 (7.82)	1.000 (31527.09)	0.058 (20.94)
4000	0.394 (20.46)	0.000 (9.42)	0.164 (9.52)	1.000 (38286.16)	0.074 (23.08)
4500	0.366 (24.62)	0.000 (11.90)	0.188 (12.57)	1.000 (43253.59)	0.082 (25.98)
5000	0.426 (28.94)	0.000 (13.46)	0.202 (14.57)	1.000 (482514.83)	0.088 (26.86)

Table A.38. Empirical power under $f \sim$ Chi-square distribution (df=3) and $g \sim$ Chi-square distribution (df=4).

n	$\tilde{\lambda_1}(A)$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.040 (6.62)	1.000 (3.75)	0.264 (2.88)	0.984 (2152.83)	1.000 (8.99)
2000	0.050 (8.74)	1.000 (4.37)	0.288 (3.72)	0.994 (4950.13)	1.000 (10.52)
2500	0.050 (10.52)	1.000 (4.46)	0.324 (3.55)	0.998 (14725.92)	1.000 (14.73)
3000	0.030 (13.89)	1.000 (5.01)	0.390 (4.65)	1.000 (22542.00)	1.000 (17.25)
3500	0.054 (15.62)	1.000 (6.83)	0.358 (4.66)	1.000 (26930.53)	1.000 (21.59)
4000	0.046 (19.72)	1.000 (8.49)	0.392 (6.79)	1.000 (31523.09)	1.000 (25.93)
4500	0.060 (23.27)	1.000 (10.25)	0.414 (8.35)	1.000 (38784.59)	1.000 (27.82)
5000	0.060 (27.98)	1.000 (12.17)	0.422 (11.74)	1.000 (43527.82)	1.000 (29.09)

Table A.39. Empirical power under $f \sim$ Chi-square distribution (df=3) and $g \sim$ Chi-square distribution (df=5).

n	$\tilde{\lambda_1}(A)$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.176 (9.03)	1.000 (3.44)	0.272 (2.58)	0.968 (2368.70)	1.000 (9.56)
2000	0.268 (18.04)	1.000 (3.47)	0.248 (2.92)	0.982 (5277.57)	1.000 (14.26)
2500	0.296 (20.53)	1.000 (4.02)	0.332 (3.35)	0.998 (17252.04)	1.000 (18.28)
3000	0.336 (24.58)	1.000 (4.24)	0.350 (5.15)	1.000 (26328.53)	1.000 (21.73)
3500	0.426 (29.05)	1.000 (4.53)	0.368 (6.26)	1.000 (30295.48)	1.000 (25.73)
4000	0.502 (32.28)	1.000 (5.92)	0.392 (7.09)	1.000 (39494.79)	1.000 (29.03)
4500	0.554 (36.17)	1.000 (6.73)	0.388 (8.55)	1.000 (45641.09)	1.000 (31.52)
5000	0.592 (41.90)	1.000 (7.27)	0.400 (10.52)	1.000 (51524.20)	1.000 (36.69)

Table A.40. Empirical power under $f \sim$ Chi-square distribution (df=3) and $g \sim$ Chi-square distribution (df=6).

n	$\tilde{\lambda_1}(A)$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.700 (8.46)	1.000 (3.48)	0.258 (3.83)	0.948 (2642.15)	1.000 (9.22)
2000	0.746 (15.74)	1.000 (3.97)	0.256 (4.62)	0.962 (5368.19)	1.000 (12.29)
2500	0.750 (23.07)	1.000 (4.43)	0.276 (5.21)	0.974 (19462.25)	1.000 (15.04)
3000	0.828 (31.55)	1.000 (4.92)	0.256 (6.10)	0.988 (25362.79)	1.000 (18.19)
3500	0.900 (36.95)	1.000 (5.83)	0.282 (7.83)	0.998 (32217.90)	1.000 (20.81)
4000	0.950 (42.90)	1.000 (6.90)	0.302 (8.82)	1.000 (38594.29)	1.000 (25.90)
4500	0.992 (47.27)	1.000 (8.28)	0.312 (9.27)	1.000 (44268.90)	1.000 (34.26)
5000	1.000 (51.78)	1.000 (10.00)	0.270 (11.42)	1.000 (52574.11)	1.000 (42.79)

Table A.41. Empirical power under $f \sim$ Chi-square distribution (df=3) and $g \sim$ chi-square distribution (df=7).

n	$\tilde{\lambda_1}(A)$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.972 (9.53)	1.000 (3.94)	0.236 (3.08)	0.972 (2493.26)	1.000 (11.04)
2000	0.992 (15.56)	1.000 (3.99)	0.286 (3.34)	0.994 (5823.83)	1.000 (14.73)
2500	0.998 (19.43)	1.000 (4.48)	0.278 (3.58)	1.000 (14268.69)	1.000 (17.26)
3000	1.000 (24.74)	1.000 (4.20)	0.282 (4.84)	1.000 (22357.00)	1.000 (19.52)
3500	1.000 (31.53)	1.000 (5.61)	0.314 (6.25)	1.000 (31345.76)	1.000 (23.25)
4000	1.000 (35.06)	1.000 (6.85)	0.306 (8.18)	1.000 (36379.59)	1.000 (27.93)
4500	0.998 (41.28)	1.000 (7.28)	0.300 (9.74)	1.000 (41426.47)	1.000 (29.91)

Table A.42. Empirical power under $f \sim$ Chi-square distribution (df=4) and $g \sim$ chi-square distribution (df=5).

n	$\tilde{\lambda_1(A)}$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.034 (8.55)	1.000 (4.14)	0.232 (3.74)	0.988 (2007.73)	1.000 (11.62)
2000	0.046 (11.20)	1.000 (5.92)	0.284 (3.91)	0.994 (6830.76)	1.000 (15.26)
2500	0.046 (14.87)	1.000 (6.85)	0.252 (4.53)	1.000 (13279.38)	1.000 (18.28)
3000	0.044 (16.21)	1.000 (7.37)	0.254 (4.85)	1.000 (21576.85)	1.000 (21.25)
3500	0.038 (18.04)	1.000 (8.09)	0.236 (5.72)	1.000 (29426.15)	1.000 (25.38)
4000	0.052 (20.93)	1.000 (8.96)	0.280 (6.92)	1.000 (35238.90)	1.000 (26.85)
4500	0.064 (25.73)	1.000 (9.93)	0.298 (7.88)	1.000 (43252.36)	1.000 (31.72)
5000	0.060 (33.93)	1.000 (11.01)	0.258 (9.46)	1.000 (51235.79)	1.000 (33.87)

Table A.43. Empirical power under $f \sim$ Gamma distribution (α,β)=(1,2) and $g \sim$ Gamma distribution (α,β)=(1,2.2).

n	$\tilde{\lambda_1(A)}$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.098 (8.81)	0.002 (3.32)	0.224 (2.36)	0.954 (1946.14)	0.600 (9.20)
2000	0.094 (15.85)	0.000 (4.12)	0.306 (3.19)	0.992 (3866.28)	0.744 (11.96)
2500	0.102 (19.98)	0.000 (4.90)	0.342 (4.48)	1.000 (8944.22)	0.826 (15.45)
3000	0.106 (23.54)	0.002 (5.73)	0.238 (5.72)	1.000 (14737.02)	0.904 (18.17)
3500	0.122 (26.47)	0.000 (7.62)	0.306 (7.11)	1.000 (21015.77)	0.920 (20.24)
4000	0.136 (28.19)	0.000 (9.52)	0.324 (8.84)	1.000 (26603.13)	0.954 (23.09)
4500	0.182 (30.20)	0.000 (10.64)	0.284 (8.90)	1.000 (35816.22)	0.994 (25.82)
5000	0.202 (34.28)	0.000 (12.27)	0.308 (9.65)	1.000 (44064.26)	0.992 (27.37)

Table A.44. Empirical power under $f \sim$ Gamma distribution (α,β)=(1,2) and $g \sim$ Gamma distribution (α,β)=(1,2.8).

n	$\tilde{\lambda_1(A)}$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.822 (8.68)	0.000 (3.08)	0.260 (2.12)	0.968 (1932.14)	1.000 (9.06)
2000	0.912 (16.46)	0.000 (3.72)	0.296 (3.85)	0.998 (3830.28)	1.000 (10.84)
2500	0.948 (21.08)	0.000 (4.35)	0.320 (4.14)	1.000 (8490.26)	1.000 (14.26)
3000	0.986 (29.36)	0.000 (5.02)	0.292 (5.66)	1.000 (15562.63)	1.000 (15.82)
3500	0.980 (34.39)	0.002 (4.61)	0.316 (6.23)	1.000 (21055.76)	1.000 (18.92)
4000	0.998 (40.69)	0.000 (6.81)	0.312 (6.58)	1.000 (26678.73)	1.000 (21.83)
4500	1.000 (41.90)	0.000 (8.53)	0.344 (7.96)	1.000 (36136.27)	1.000 (24.27)
5000	1.000 (48.73)	0.004 (10.48)	0.322 (9.03)	1.000 (43173.00)	1.000 (28.96)

Table A.45. Empirical power under $f \sim$ Gamma distribution (α,β)=(1,2) and $g \sim$ Gamma distribution (α,β)=(1,2.5).

n	$\lambda_1(\tilde{A})$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.198 (9.27)	0.378 (2.84)	0.294 (2.89)	0.950 (2384.85)	1.000 (9.68)
2000	0.234 (14.52)	0.470 (3.53)	0.294 (3.57)	0.984 (4857.19)	1.000 (13.48)
2500	0.244 (19.81)	0.594 (3.16)	0.262 (4.86)	1.000 (9533.64)	1.000 (15.65)
3000	0.466 (24.95)	0.624 (4.34)	0.312 (5.93)	1.000 (16235.72)	1.000 (18.67)
3500	0.592 (29.53)	0.738 (4.59)	0.360 (6.15)	1.000 (23427.99)	1.000 (21.27)
4000	0.646 (34.26)	0.830 (6.86)	0.330 (7.27)	1.000 (29058.24)	1.000 (24.14)
4500	0.694 (39.64)	0.872 (6.07)	0.324 (8.94)	1.000 (37976.23)	1.000 (27.79)
5000	0.802 (45.26)	0.888 (6.13)	0.338 (11.43)	1.000 (43064.46)	1.000 (29.26)

Table A.46. Empirical power under $f \sim$ Gamma distribution (α,β)=(1,2) and $g \sim$ Gamma distribution (α,β)=(1,3).

n	$\lambda_1(\tilde{A})$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.998 (9.01)	1.000 (3.42)	0.244 (2.56)	0.974 (2465.18)	1.000 (9.40)
2000	1.000 (15.72)	1.000 (4.04)	0.268 (3.37)	0.994 (4264.27)	1.000 (13.82)
2500	1.000 (18.79)	1.000 (4.18)	0.282 (3.27)	0.998 (7826.74)	1.000 (16.26)
3000	1.000 (23.83)	1.000 (4.53)	0.286 (4.18)	1.000 (14266.27)	1.000 (19.07)
3500	1.000 (28.26)	1.000 (4.92)	0.344 (5.74)	1.000 (24256.53)	1.000 (22.67)
4000	1.000 (32.17)	1.000 (6.62)	0.298 (6.84)	1.000 (35435.90)	1.000 (25.58)
4500	1.000 (36.81)	1.000 (6.88)	0.302 (8.33)	1.000 (41252.67)	1.000 (28.95)
5000	1.000 (41.90)	1.000 (7.33)	0.342 (9.46)	1.000 (49228.93)	1.000 (31.58)

Table A.47. Empirical power under $f \sim$ Gamma distribution (α,β)=(1.4,2.5) and $g \sim$ Gamma distribution (α,β)=(1,2.5).

n	$\lambda_1(\tilde{A})$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.804 (9.12)	1.000 (3.53)	0.276 (2.67)	0.988 (2325.14)	1.000 (9.51)
2000	0.940 (14.27)	1.000 (3.70)	0.260 (3.05)	0.996 (4687.39)	1.000 (11.82)
2500	0.992 (17.19)	1.000 (4.82)	0.277 (3.77)	0.998 (7575.57)	1.000 (15.18)
3000	0.998 (21.42)	1.000 (5.90)	0.336 (4.57)	1.000 (13256.50)	1.000 (18.63)
3500	1.000 (26.38)	1.000 (5.72)	0.294 (5.93)	1.000 (19734.96)	1.000 (21.74)
4000	1.000 (31.57)	1.000 (6.68)	0.332 (6.10)	1.000 (24567.14)	1.000 (23.36)
4500	1.000 (36.88)	1.000 (6.49)	0.334 (8.89)	1.000 (35627.26)	1.000 (25.12)
5000	1.000 (42.89)	1.000 (7.27)	0.340 (8.41)	1.000 (47383.16)	1.000 (28.04)

Table A.48. Empirical power under $f \sim$ Gamma distribution (α,β)=(1,1.5) and $g \sim$ Gamma distribution (α,β)=(1,1.8).

n	$\tilde{\lambda_1}(A)$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.204 (7.87)	1.000 (3.28)	0.284 (2.42)	0.978 (1949.63)	1.000 (7.26)
2000	0.358 (15.21)	1.000 (3.64)	0.300 (2.99)	0.992 (3899.47)	1.000 (11.89)
2500	0.484 (20.50)	1.000 (4.99)	0.322 (3.08)	0.996 (8253.26)	1.000 (14.62)
3000	0.536 (25.87)	1.000 (5.39)	0.330 (4.12)	1.000 (15868.30)	1.000 (19.71)
3500	0.568 (31.59)	1.000 (6.67)	0.320 (4.58)	1.000 (23280.06)	1.000 (23.89)
4000	0.548 (37.63)	1.000 (7.42)	0.352 (5.56)	1.000 (29362.95)	1.000 (27.21)
4500	0.782 (41.27)	1.000 (7.90)	0.368 (7.49)	1.000 (34622.47)	1.000 (31.35)
5000	0.884 (45.26)	1.000 (8.28)	0.412 (8.43)	1.000 (48338.40)	1.000 (38.83)

Table A.49. Empirical power under $f \sim$ Gamma distribution (α,β)=(1,1.5) and $g \sim$ Gamma distribution (α,β)=(0.9,1.5).

n	$\tilde{\lambda_1}(A)$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.146 (6.72)	0.000 (2.94)	0.312 (3.01)	0.944 (1772.37)	0.104 (8.81)
2000	0.162 (11.62)	0.000 (3.63)	0.340 (3.85)	0.984 (4373.46)	0.144 (11.75)
2500	0.230 (15.25)	0.000 (4.47)	0.374 (4.54)	0.998 (9428.48)	0.130 (14.55)
3000	0.246 (18.72)	0.000 (5.79)	0.316 (5.89)	0.998 (17371.56)	0.168 (18.52)
3500	0.294 (21.09)	0.000 (6.28)	0.346 (7.11)	1.000 (26236.94)	0.196 (21.84)
4000	0.302 (26.48)	0.002 (7.84)	0.324 (6.53)	1.000 (33526.27)	0.206 (26.76)
4500	0.362 (31.77)	0.000 (9.72)	0.288 (8.21)	1.000 (40285.38)	0.196 (30.64)
5000	0.388 (38.80)	0.000 (11.58)	0.362 (9.57)	1.000 (49062.03)	0.224 (36.23)

Table A.50. Empirical power under $f \sim$ Gamma distribution (α,β)=(0,1) and $g \sim$ Cauchy distribution (α,β)=(0,1.2).

n	$\tilde{\lambda_1}(A)$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	0.998 (8.65)	0.000 (3.76)	0.294 (3.16)	0.958 (1830.72)	0.058 (10.00)
2000	0.998 (10.16)	0.000 (4.11)	0.358 (4.21)	0.994 (4729.14)	0.036 (11.94)
2500	1.000 (13.74)	0.000 (5.02)	0.346 (6.92)	1.000 (10572.19)	0.058 (14.79)
3000	1.000 (16.03)	0.000 (6.68)	0.368 (8.11)	1.000 (15232.95)	0.056 (17.98)
3500	1.000 (19.21)	0.000 (8.53)	0.374 (7.90)	1.000 (22926.21)	0.052 (20.04)
4000	1.000 (23.90)	0.000 (10.05)	0.402 (8.81)	1.000 (29580.98)	0.054 (23.57)
4500	1.000 (26.84)	0.002 (13.72)	0.444 (10.73)	1.000 (37018.65)	0.034 (28.26)
5000	1.000 (33.17)	0.000 (16.83)	0.428 (10.51)	1.000 (46749.99)	0.050 (29.70)

Table A.51. Empirical power under $f \sim$ Gamma distribution (α,β) =(2,4) and $g \sim$ Cauchy distribution (α,β)=(2.5,5).

n	$\tilde{\lambda_1(A)}$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
1500	1.000 (9.33)	0.002 (3.74)	0.248 (2.88)	0.978 (1799.51)	0.434 (9.72)
2000	1.000 (18.17)	0.022 (4.10)	0.280 (2.95)	0.984 (4254.89)	0.504 (11.72)
2500	1.000 (20.72)	0.036 (4.27)	0.310 (3.34)	1.000 (11728.87)	0.622 (14.69)
3000	1.000 (27.16)	0.026 (5.21)	0.322 (6.76)	1.000 (16822.37)	0.662 (18.52)
3500	1.000 (32.16)	0.050 (6.33)	0.282 (9.31)	1.000 (21868.69)	0.764 (20.78)
4000	1.000 (36.92)	0.044 (6.39)	0.318 (7.94)	1.000 (28636.65)	0.808 (23.23)
4500	1.000 (47.72)	0.038 (9.12)	0.388 (8.68)	1.000 (34681.30)	0.844 (25.98)
5000	1.000 (52.84)	0.064 (11.53)	0.392 (10.63)	1.000 (45739.49)	0.870 (27.67)

Table A.52. Empirical power under $f \sim$ Chi-square distribution (df=2) and $g \sim$ F-distribution (α,β)=(2,4).

n	$\tilde{\lambda_1(A)}$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
150	0.688 (0.41)	0.428 (1.84)	0.654 (346.31)	0.875 (292.43)	0.714 (0.72)
300	0.782 (0.80)	0.694 (2.29)	0.732 (1210.25)	0.928 (1062.89)	0.803 (1.21)
450	0.874 (0.96)	0.840 (2.07)	0.858 (2724.03)	0.945 (1306.07)	0.859 (1.36)
600	0.952 (1.25)	0.924 (3.92)	0.934 (4738.86)	1.000 (4296.68)	0.939 (1.99)
750	0.982 (2.48)	0.980 (3.63)	0.992 (7189.28)	1.000 (6274.58)	0.994 (2.21)
900	1.000 (2.86)	1.000 (3.50)	1.000 (10598.49)	1.000 (9528.81)	1.000 (2.89)

Table A.53. Empirical power under $f \sim$ Gamma distribution (α,β)=(10,5) and $g \sim$ Normal-distribution (μ,σ)=(50,16).

n	$\tilde{\lambda_1(A)}$ (time elapsed(s))	KS (time elapsed(s))	MMD (time elapsed(s))	AD (time elapsed(s))	MWU (time elapsed(s))
150	0.624 (0.52)	0.400 (2.13)	0.724 (322.27)	0.818 (264.82)	0.684 (1.33)
300	0.736 (0.78)	0.686 (2.72)	0.822 (1312.60)	0.930 (1193.38)	0.752 (1.42)
450	0.898 (0.99)	0.794 (3.13)	0.892 (2577.62)	0.972 (2296.80)	0.884 (2.02)
600	0.966 (1.42)	0.910 (3.40)	0.974 (4595.46)	0.998 (4146.25)	0.946 (2.45)
750	0.980 (2.36)	0.952 (3.78)	0.992 (7543.28)	1.000 (6962.38)	0.992 (2.88)
900	0.998 (2.91)	1.000 (3.94)	1.000 (12170.49)	1.000 (11447.10)	0.998 (3.25)