Statistics and Probability Letters 216 (2025) 110278

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Universally consistent K-sample tests via dependence measures

Sambit Panda [a,1], Cencheng Shen [b,1], Ronan Perry [a], Jelle Zorn [c], Antoine Lutz [c], Carey E. Priebe [d], Joshua T. Vogelstein [a,e,*]

[a] Department of Biomedical Engineering, Johns Hopkins University, MD, USA
[b] Department of Applied Economics and Statistics, University of Delaware, DE, USA
[c] Lyon Neuroscience Research Centre, Universite Claude Bernard Lyon 1, Lyon, France
[d] Department of Applied Mathematics and Statistics, Johns Hopkins University, MD, USA
[e] Center for Imaging Science and Kavli Neuroscience Discovery Institute, Johns Hopkins University, MD, USA

ARTICLE INFO

Keywords:
K-sample testing
Testing independence
Dependence measure

ABSTRACT

The K-sample testing problem involves determining whether K groups of data points are each drawn from the same distribution. Analysis of variance is arguably the most classical method to test mean differences, along with several recent methods to test distributional differences. In this paper, we demonstrate the existence of a transformation that allows K-sample testing to be carried out using any dependence measure. Consequently, universally consistent K-sample testing can be achieved using a universally consistent dependence measure, such as distance correlation and the Hilbert–Schmidt independence criterion. This enables a wide range of dependence measures to be easily applied to K-sample testing.

1. Introduction

Given two datasets $\{u_i^{(1)} \in \mathbb{R}^p, i = 1, \ldots, n_1\}$ and $\{u_j^{(2)} \in \mathbb{R}^p, j = 1, \ldots, n_2\}$, assume each $u_i^{(1)}$ is sampled independently and identically (i.i.d.) from F_{U_1}, and each $u_j^{(2)}$ is sampled i.i.d. from F_{U_2}. Also, assume that each pair $(u_i^{(1)}, u_j^{(1)})$ is independent for any (i, j). The classical two-sample testing problem tests whether the two datasets were sampled from the same distribution, stated as:

$$H_0 : F_{U_1} = F_{U_2},$$
$$H_A : F_{U_1} \neq F_{U_2}.$$

The K-sample testing problem is a generalization of the above. Let $\{u_i^{(k)} \overset{i.i.d.}{\sim} F_{U_k} \in \mathbb{R}^p, i = 1, \ldots, n_k\}$ for $k = 1, \ldots, K$, and assume $(u_i^{(s)}, u_j^{(t)})$ is independent for any (i, j, s, t). We aim to test:

$$H_0 : F_{U_1} = F_2 = \cdots = F_{U_K},$$
$$H_A : \exists\, s \neq t \text{ s.t. } F_{U_s} \neq F_{U_t}.$$

Student's t-test and its multivariate generalization, Hotelling's T^2, are traditionally used for two-sample testing, while analysis of variance (`Anova`) or multivariate analysis of variance (`Manova`) are conventional choices for K-sample tests. These tests, however, only aim to test mean differences, and do not perform well for non-Gaussian data beyond their parametric assumptions (Warne, 2014). To address this, several nonparametric statistics have been developed to test distributional differences, such as

* Corresponding author at: Department of Biomedical Engineering, Johns Hopkins University, MD, USA.
E-mail address: jovo@jhu.edu (J.T. Vogelstein).

[1] Sambit Panda and Cencheng Shen contribute equally to this work.

https://doi.org/10.1016/j.spl.2024.110278
Received 1 June 2024; Received in revised form 14 September 2024; Accepted 16 September 2024
Available online 19 September 2024

Energy (Székely and Rizzo, 2013b) and maximal mean discrepancy (Mmd) (Gretton et al., 2012) for two-sample tests, and multivariate Heller–Heller–Gorfine (Heller et al., 2016) and distance components (Disco) (Rizzo and Székely, 2010) for K-sample tests.

A closely related and popular problem in statistics is testing independence. Given $x_i \in \mathbb{R}^p$ and $y_i \in \mathbb{R}^q$, and n samples of $(x_i, y_i) \overset{iid}{\sim} F_{XY}$, the independence hypothesis is stated as:

$$H_0 : F_{XY} = F_X F_Y,$$
$$H_A : F_{XY} \neq F_X F_Y.$$

Traditional Pearson's correlation (Pearson, 1895) is popular but limited to detecting linear dependence. Many non-parametric methods have been proposed recently, such as distance covariance (Dcov) (Székely et al., 2007; Székely and Rizzo, 2013a), Hilbert–Schmidt independence criterion (Hsic) (Gretton et al., 2005; Gretton and Györfi, 2010; Bounliphone et al., 2016), multiscale graph correlation (Mgc) (Vogelstein et al., 2019; Shen et al., 2020), among many others (Heller et al., 2012; Pan et al., 2020; Zhou et al., 2024; Panda et al., 2024b).

These recent dependence measures are universally consistent under mild assumptions, such as finite moments. Specifically, the sample statistic converges to a population statistic, which equals 0 if and only if X and Y are independent. Therefore, these measures can detect any type of relationship given a sufficiently large sample size, and the testing power increases to 1 as n increases, regardless of whether the underlying relationship is linear or nonlinear. These universally consistent dependence measures have found applications for various inference tasks, such as feature screening (Li et al., 2012; Zhong and Zhu, 2015; Shen et al., 2024e; Shen and Dong, 2024), time-series analysis (Zhou, 2012; Shen et al., 2024b), conditional independence (Fukumizu et al., 2007; Székely and Rizzo, 2014; Wang et al., 2015), and graph testing (Lee et al., 2019; Shen et al., 2024a).

In this manuscript, we establish a fundamental connection between K-sample testing and independence testing: there exists a transformation of the given data such that the K-sample testing problem is converted to the independence testing problem on the transformed data. As a result, any universally consistent dependence measure can be used to achieve universally consistent K-sample testing. Moreover, the proposed sample transformation allows previously established two-sample statistics, such as Energy and Mmd, to be equivalent to the corresponding dependence measures, Dcov and Hsic, respectively; and the more general K-sample Disco equals a bootstrap version of Dcov. Finally, we used simulated data to verify the validity, consistency, and finite-sample testing power of several popular dependence measures for K-sample testing. Theorem proofs and additional simulations are provided in the supplementary material. The code is available in the Hyppo statistical package (Panda et al., 2024a) and on GitHub.[2]

2. Method and theory

In this section, we first review Dcov and Hsic, which are the foundational blocks of universally consistent dependence measures. Next, we introduce the population transformation that converts K-sample testing to independence testing in the random variable setting, enabling any consistent dependence measure to achieve consistent K-sample testing. We then proceed to the sample method and prove the sample equivalence between Dcov and Energy, and between Hsic and Mmd. Throughout this section, we assume all distributions have finite moments, and all proofs are provided in the appendix.

2.1. Review of dependence measures

Denote the paired sample data as $(\mathbf{X}, \mathbf{Y}) = \{(x_i, y_i) \in \mathbb{R}^{p+q}, i = 1, \ldots, n\}$, where each sample pair (x_i, y_i) is assumed to be i.i.d. as F_{XY} with finite moments. Given a distance metric $d(\cdot,\cdot) : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$, such as the Euclidean metric, denote $\mathbf{D}^{\mathbf{X}}$ and $\mathbf{D}^{\mathbf{Y}}$ as the $n \times n$ distance matrices for $\mathbf{X}$ and $\mathbf{Y}$, respectively. Define $\mathbf{H} = \vec{I} - \frac{1}{n}\vec{J}$ as an $n \times n$ centering matrix, where $\vec{I}$ is the identity matrix and $\vec{J}$ is the matrix of ones. Then the sample distance covariance (Dcov) and distance correlation (Dcor) can be computed by:

$$\texttt{Dcov}_n(\mathbf{X}, \mathbf{Y}) = \frac{1}{n^2} trace(\mathbf{H}\mathbf{D}^{\mathbf{X}}\mathbf{H}\mathbf{H}\mathbf{D}^{\mathbf{Y}}\mathbf{H}),$$
$$\texttt{Dcor}_n(\mathbf{X}, \mathbf{Y}) = \frac{\texttt{Dcov}_n(\mathbf{X}, \mathbf{Y})}{\sqrt{\texttt{Dcov}_n(\mathbf{X}, \mathbf{X})\texttt{Dcov}_n(\mathbf{Y}, \mathbf{Y})}}.$$

By default, distance correlation utilizes the Euclidean distance as its metric, but it can be any metric of strong negative type (Lyons, 2013), or a characteristic kernel upon a kernel to metric transformation (Sejdinovic et al., 2013; Shen and Vogelstein, 2021). Moreover, by replacing the distance metric $d(\cdot,\cdot)$ with a kernel measure $k(\cdot,\cdot)$, such that $\mathbf{D}^{\mathbf{X}}$ and $\mathbf{D}^{\mathbf{Y}}$ become the corresponding kernel matrices, Dcov becomes Hsic.

[2] https://hyppo.neurodata.io/

2.2. Population transformation

Theorem 1. *Given K random variables $(U_1, U_2, \ldots, U_K)$. Let $V \in \mathbb{R}^K$ be the multinomial distribution of probability $(\pi_1, \pi_2, \ldots, \pi_K)$, where $\pi_k \in (0,1)$ and $\sum_{k=1}^{K} \pi_k = 1$. Let U be the following mixture distribution:*

$$U = \sum_{k=1}^{K} 1(V_k = 1) U_k,$$

where V_k denotes the kth dimension of V. Then, $F_{UV} = F_U F_V$ if and only if $F_{U_1} = F_{U_2} = \cdots = F_{U_K}$.

Therefore, the proposed transformation from $(U_1, U_2, \ldots, U_K) \to (U, V)$ converts K-sample testing on $(U_1, U_2, \ldots, U_K)$ to independence testing on (U, V), leading to the consistency of dependence measures for K-sample testing.

Corollary 1. *Suppose $\tau(\cdot,\cdot)$ is a universally consistent dependence measure such that $\tau(X, Y) = 0$ if and only if X and Y are independent. Then, via the proposed transformation in Theorem 1, $\tau(U, V)$ is universally consistent for the K-sample test on U_k, i.e., $\tau(U, V) = 0$ if and only if $F_{U_1} = F_{U_2} = \cdots = F_{U_K}$.*

Note that the probabilities $\{\pi_k\}$ can be easily chosen based on sample size, as described in the sample method subsection below. However, the transformation and the resulting consistency apply to any $\{\pi_k\}$, as long as all probabilities are positive, ensuring that each random variable U_k has a positive probability of appearing in the mixture U. For example, if we use `Dcov` as the statistic, the value of $\texttt{Dcov}(U, V)$ may differ for different choices of $\{\pi_k\}$ when some distributions are different. Nonetheless, $\texttt{Dcov}(U, V) = 0$ if and only if all distributions are the same, which holds for any choice of $\{\pi_k\}$.

2.3. Sample method

Given the sample data and a sample dependence measure $\tau_n(\cdot,\cdot)$, we can carry out the K-sample testing as follows:

- **Input**: For each $k = 1, \ldots, K$, the sample data $\mathbf{U}_k = [u_1^{(k)}, \ldots, u_{n_k}^{(k)}]^T \in \mathbb{R}^{n_k \times p}$; a given sample dependence measure $\tau_n(\cdot,\cdot) : \mathbb{R}^{n \times p} \times \mathbb{R}^{n \times K} \to \mathbb{R}$.
- **Sample Transformation**: Let $n = \sum_{k=1}^{K} n_k$, we set $\mathbf{U} \in \mathbb{R}^{n \times p}$ as the row-concatenation of all data matrices, and $\mathbf{V} \in \mathbb{R}^{n \times K}$ as a one-hot encoding of the data source label. That is,

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_1 \\ \vdots \\ \mathbf{U}_K \end{bmatrix} \in \mathbb{R}^{n \times p},$$

$$\mathbf{V} = \begin{bmatrix} \vec{1}_{n_1 \times 1} & \vec{0}_{n_1 \times 1} & \cdots & \vec{0}_{n_1 \times 1} \\ \vec{0}_{n_2 \times 1} & \vec{1}_{n_2 \times 1} & \cdots & \vec{0}_{n_2 \times 1} \\ \vdots & \vdots & \ddots & \vdots \\ \vec{0}_{n_K \times 1} & \vec{0}_{n_K \times 1} & \cdots & \vec{1}_{n_K \times 1} \end{bmatrix} \in \mathbb{R}^{n \times K},$$

  where $\vec{1}_{n_1 \times 1}$ is the vector of ones of length n_1.
- **Dependence Measure**: Compute $\tau_n(\mathbf{U}, \mathbf{V})$.
- **Permutation Test**: Compute the p-value via a standard permutation test, i.e.,

$$\text{pval} = \frac{1}{R} \sum_{r=1}^{R} 1\{\tau_n(\mathbf{U}, \text{perm}_r(\mathbf{V})) > \tau_n(\mathbf{U}, \mathbf{V})\}$$

  where $\text{perm}_r()$ represents a random permutation of size n (which is different and randomized for each r), and R is the number of permutations.
- **Output**: The sample statistic $\tau_n(\mathbf{U}, \mathbf{V})$ and its p-value.

Here, the sample matrix pair $(\mathbf{U}, \mathbf{V})$ can be viewed as the sample realization of the population transformation (U, V), where the mixture probability $\{\pi_k = \frac{n_k}{n}, k = 1, \ldots, K\}$. The one-hot encoding scheme has been a fundamental technique in neural networks and machine learning (Bishop, 1995; Murphy, 2012) and has recently been applied in graph embedding (Shen et al., 2023, 2024c,d). For the choice of dependence measure, $\tau_n(\mathbf{U}, \mathbf{V})$ can be any aforementioned sample dependence measure, such as `Dcov`, `Hsic`, `Mgc`, etc. While the permutation test is used here, one could use a faster testing procedure, such as the chi-square test via an unbiased test statistic (Shen et al., 2022).

Because $\mathbf{V}$ is categorical, the distance between any two rows of $\mathbf{V}$ can only take two values. Specifically, $d(\mathbf{V}(i,:), \mathbf{V}(j,:))$ is either 0 when $\mathbf{V}(i,:) = \mathbf{V}(j,:)$ or $\sqrt{2}$ when $\mathbf{V}(i,:) \neq \mathbf{V}(j,:)$ under Euclidean distance. The former occurs when the ith and jth sample data come from the same group, and the latter occurs when they come from different groups. We use β to denote the maximum distance minus the minimum distance within the distance matrix of $\mathbf{V}$. As the first and last observations in $\mathbf{V}$ always come from different groups based on our construction, we can conveniently let $\beta = d(\mathbf{V}(1,:), \mathbf{V}(n,:))$ in this case, where $\mathbf{V}(1,:)$ represents the first row and $\mathbf{V}(n,:)$ represents the last row of the matrix.

2.4. Sample properties

Given the above sample transformation, the sample distance covariance for $(\mathbf{U}, \mathbf{V})$ can be proved to be exactly the same as the sample energy statistic for $(\mathbf{U}_1, \mathbf{U}_2)$, up to a scalar constant.

Theorem 2. *Assume a translation-invariant metric $d(\cdot,\cdot)$ is used, and denote $\beta = d(\mathbf{V}(1,:), \mathbf{V}(n,:))$. It follows that*

$$\texttt{Dcov}_n(\mathbf{U}, \mathbf{V}) = \frac{2n_1^2 n_2^2 \beta}{n^4} \cdot \texttt{Energy}_{n_1,n_2}(\mathbf{U}_1, \mathbf{U}_2),$$

Under the permutation test, distance covariance, distance correlation, and energy statistic have the same testing p-value.

By default, both Dcov and Energy use Euclidean distance, which is translation-invariant. Moreover, due to the existing transformation between distance and kernel, Hsic is also equivalent to Mmd.

Theorem 3. *Assume a translation-invariant kernel $k(\cdot,\cdot)$ is used, and denote $\beta = k(\mathbf{V}(1,:), \mathbf{V}(n,:)) - k(\mathbf{V}(1,:), \mathbf{V}(1,:))$. It follows that*

$$\texttt{Hsic}_n(\mathbf{U}, \mathbf{V}) = \frac{2n_1^2 n_2^2 \beta}{n^4} \cdot \texttt{Mmd}_{n_1,n_2}(\mathbf{U}_1, \mathbf{U}_2),$$

Under the permutation test, Hilbert–Schmidt independence criterion and maximum mean discrepancy have the same testing p-value.

Lastly, in the case of general K-sample testing, the sample distance covariance for $(\mathbf{U}, \mathbf{V})$ is a weighted summation of pairwise two-sample energy statistics. This coincides with the K-sample Disco statistic if the data sources are equally weighted.

Theorem 4. *Assume a translation-invariant metric $d(\cdot,\cdot)$ is used, and denote $\beta = d(\mathbf{V}(1,:), \mathbf{V}(n,:))$. It follows that*

$$\texttt{Dcov}_n(\mathbf{U}, \mathbf{V}) = \beta \sum_{1 \le s < t \le K} \left\{ \frac{n(n_s + n_t) - \sum_{l=1}^{K} n_l^2}{n^4} \cdot n_s n_t \texttt{Energy}_{n_s,n_t}(\mathbf{U}_s, \mathbf{U}_t) \right\}.$$

Moreover, sample distance covariance is equivalent to the sample Disco *statistic when $n_1 = n_2 = \cdots = n_k$, in which case*

$$\texttt{Dcov}_n(\mathbf{U}, \mathbf{V}) = \frac{2\beta}{nK} \texttt{Disco}(\{\mathbf{U}_k\}).$$

Therefore, if we consider a bootstrap resampling of the given data $\{\mathbf{U}_k\}$, such that the bootstrap samples consist of the same number of observations per data source, and let $\mathbf{U}$ and $\mathbf{V}$ be the concatenated bootstrap samples, this will enforce $n_1 = n_2 = \cdots = n_K$. Thus, distance covariance for K-sample testing is equivalent to Disco on equally-weighted bootstrap samples.

It is worth noting that Proposition 6 and Corollary 5 of Edelmann and Goeman (2022) also establish the equivalence between Hsic and Mmd, which aligns with our Theorem 3 for a kernel choice where $\beta = 1$. Their work focused on discrete kernels for categorical data and found that Hsic coincides with Mmd in this context. In contrast, our paper employs one-hot encoding to transform the K-sample problem into an independence test, with $\mathbf{V}$ always being discrete. Interestingly, despite starting from different perspectives and following different procedures, we arrived at the same equivalence.

3. Simulations

We aim to verify the validity and consistency of using dependence measures for two-sample tests. Specifically, we consider three univariate settings and compare Anova, distance correlation, and Hsic in each case. Note that more numerical comparisons involving other dependence measures and non-Gaussian simulations are provided in appendix Figure F1.

We set up two random variables U_1 and U_2 and the corresponding parameters as follows:

1. **Sample Size Difference**: Both U_1 and U_2 are standard normal. $n_1 = 100$, and $n_2 = 20, 40, \ldots, 200$.
2. **Mean Difference**: U_1 is standard normal, and $U_2 \sim \text{Normal}(c, 1)$, where $c = 0, 0.05, 0.1, \ldots, 0.5$. $n_1 = 100$, and $n_2 = 200$.
3. **Variance Difference**: U_1 is standard normal, and $U_2 \sim \text{Normal}(0, 1 + c)$, where $c = 0, 0.1, \ldots, 1$. $n_1 = 100$, and $n_2 = 200$.

We generate sample data $\mathbf{U}_1$ and $\mathbf{U}_2$ accordingly, and compute the p-value for each method. This is repeated for 1000 replicates, and we compute the testing power of each method by checking how often the p-value is lower than at the type 1 error level 0.05.

Fig. 1 shows the testing power for each simulation. In the first simulation (left panel), there is no distribution difference; only the sample sizes are different, and all methods have a testing power of about 0.05. In the second simulation (center panel), the mean difference is detected by all three methods, with Anova and Dcov showing the best power, followed by Hsic. In the last simulation (right panel), Anova has little power, while Hsic performs the best, followed by Dcov.

Overall, the first simulation demonstrates the validity of all methods, while the second and third simulations demonstrate the consistency of dependence measures in testing distributional differences. As Anova is designed to detect mean differences in Gaussian settings, it works for the second simulation but not the third. We also notice that Dcov is better at detecting mean differences, while Hsic is better at detecting variance differences, which can be attributed to the fact that Dcov has better finite-sample power for detecting linear dependence, while Hsic is better at detecting nonlinear dependence. We note that while universally consistent dependence measures are guaranteed to achieve perfect testing power given a sufficiently large sample size, different dependence measures can excel at detecting different distributions, requiring less sample size to achieve power 1.

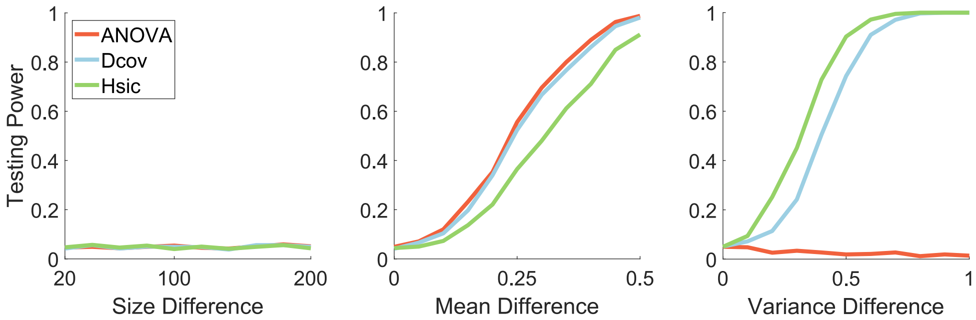

Fig. 1. The figure compares the testing power of Anova, Dcov, and Hsic for three different Gaussian-simulated sample datasets.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work is supported by the Defense Advanced Research Projects Agency (DARPA) Lifelong Learning Machines program through contract FA8650-18-2-7834, the National Institutes of Health awards RO1MH120482 and T32GM119998, and the National Science Foundation award DMS-1921310 and DMS-2113099. The authors would like to thank the editor, two anonymous reviewers, Dr. Russell Lyons, Dr. Minh Tang, Mr. Ronak Mehta, Mr. Eric W. Bridgeford, and the rest of the NeuroData group at Johns Hopkins University for valuable feedback.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.spl.2024.110278.

References

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press.
Bounliphone, W., Belilovsky, E., Blaschko, M.B., Antonoglou, I., Gretton, A., 2016. A test of relative similarity for model selection in generative models. In: International Conference on Learning Representations.
Edelmann, D., Goeman, J., 2022. A regression perspective on generalized distance covariance and the Hilbert–Schmidt independence criterion. Statist. Sci. 37, 562–579.
Fukumizu, K., Gretton, A., Sun, X., Schölkopf, B., 2007. Kernel measures of conditional dependence. In: Platt, J., Koller, D., Singer, Y., Roweis, S. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc., pp. 489–496.
Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B., Smola, A., 2012. A kernel two-sample test. J. Mach. Learn. Res. 13, 723–773.
Gretton, A., Györfi, L., 2010. Consistent nonparametric tests of independence. J. Mach. Learn. Res. 11, 1391–1423.
Gretton, A., Herbrich, R., Smola, A., Bousquet, O., Schölkopf, B., 2005. Kernel methods for measuring independence. J. Mach. Learn. Res. 6, 2075–2129.
Heller, R., Heller, Y., Gorfine, M., 2012. A consistent multivariate test of association based on ranks of distances. Biometrika 100, 503–510.
Heller, R., Heller, Y., Kaufman, S., Brill, B., Gorfine, M., 2016. Consistent distribution-free k-sample and independence tests for univariate random variables. J. Mach. Learn. Res. 17, 978–1031.
Lee, Y., Shen, C., Priebe, C.E., Vogelstein, J.T., 2019. Network dependence testing via diffusion maps and distance-based correlations. Biometrika 106, 857–873.
Li, R., Zhong, W., Zhu, L., 2012. Feature screening via distance correlation learning. J. Amer. Statist. Assoc. 107, 1129–1139.
Lyons, R., 2013. Distance covariance in metric spaces. Ann. Probab. 41, 3284–3305.
Murphy, K.P., 2012. Machine Learning: A Probabilistic Perspective. MIT Press.
Pan, W., Wang, X., Zhang, H., Zhu, H., Zhu, J., 2020. Ball covariance: A generic measure of dependence in banach space. J. Amer. Statist. Assoc. 115, 307–317.
Panda, S., Palaniappan, S., Xiong, J., Bridgeford, E.W., Mehta, R., Shen, C., Vogelstein, J., 2024a. hyppo: A multivariate hypothesis testing python package. arXiv preprint arXiv:1907.02088.
Panda, S., Shen, C., Vogelstein, J.T., 2024b. Learning interpretable characteristic kernels via decision forests. arXiv preprint arXiv:1812.00029.
Pearson, K., 1895. Vii. note on regression and inheritance in the case of two parents. In: Proceedings of the Royal Society of London, Vol. 58. pp. 240–242.
Rizzo, M.L., Székely, G.J., 2010. Disco analysis: A nonparametric extension of analysis of variance. Ann. Appl. Stat. 4, 1034–1055.
Sejdinovic, D., Sriperumbudur, B., Gretton, A., Fukumizu, K., 2013. Equivalence of distance-based and rkhs-based statistics in hypothesis testing. Ann. Statist. 41, 2263–2291.
Shen, C., Arroyo, J., Xiong, J., Vogelstein, J.T., 2024a. Community correlations and testing independence between binary graphs. arXiv preprint arXiv:1906.03661.
Shen, C., Chung, J., Mehta, R., Xu, T., Vogelstein, J.T., 2024b. Independence testing for temporal data. Trans. Mach. Learn. Res. URL: https://openreview.net/forum?id=jv1aPQINc4.
Shen, C., Dong, Y., 2024. High-dimensional independence testing via maximum and average distance correlations. arXiv preprint arXiv:2001.01095.
Shen, C., Larson, J., Trinh, H., Qin, X., Park, Y., Priebe, C.E., 2024c. Discovering communication pattern shifts in large-scale labeled networks using encoder embedding and vertex dynamics. IEEE Trans. Netw. Sci. Eng. 11, 2100–2109.
Shen, C., Panda, S., Vogelstein, J.T., 2022. The chi-square test of distance correlation. J. Comput. Graph. Statist. 31, 254–262.
Shen, C., Priebe, C.E., Larson, J., Trinh, H., 2024d. Synergistic graph fusion via encoder embedding. Inform. Sci. 678, 120912.
Shen, C., Priebe, C.E., Vogelstein, J.T., 2020. From distance correlation to multiscale graph correlation. J. Amer. Statist. Assoc. 115, 280–291.

Shen, C., Vogelstein, J.T., 2021. The exact equivalence of distance and kernel methods in hypothesis testing. AStA Adv. Stat. Anal. 105, 385–403.
Shen, C., Wang, S., Badea, A., Priebe, C.E., Vogelstein, J.T., 2024e. Discovering the signal subgraph: An iterative screening approach on graphs. Pattern Recognit. Lett. 184, 97–102.
Shen, C., Wang, Q., Priebe, C.E., 2023. One-hot graph encoder embedding. IEEE Trans. Pattern Anal. Mach. Intell. 45, 7933–7938.
Székely, G.J., Rizzo, M.L., 2013a. The distance correlation t-test of independence in high dimension. J. Multivariate Anal. 117, 193–213.
Székely, G.J., Rizzo, M.L., 2013b. Energy statistics: A class of statistics based on distances. J. Statist. Plann. Inference 143, 1249–1272.
Székely, G.J., Rizzo, M.L., 2014. Partial distance correlation with methods for dissimilarities. Ann. Statist. 42, 2382–2412.
Székely, G.J., Rizzo, M.L., Bakirov, N.K., 2007. Measuring and testing dependence by correlation of distances. Ann. Stat. 35, 2769–2794.
Vogelstein, J.T., Bridgeford, E.W., Wang, Q., Priebe, C.E., Maggioni, M., Shen, C., 2019. Discovering and deciphering relationships across disparate data modalities. eLife 8, e41690.
Wang, X., Pan, W., Hu, W., Tian, Y., Zhang, H., 2015. Conditional distance correlation. J. Amer. Statist. Assoc. 110, 1726–1734.
Warne, R., 2014. A primer on multivariate analysis of variance (manova) for behavioral scientists. Pract. Assess. Res. Eval. 19 (17).
Zhong, W., Zhu, L., 2015. An iterative approach to distance correlation-based sure independence screening. J. Stat. Comput. Simul. 85, 2331–2345.
Zhou, Z., 2012. Measuring nonlinear dependence in time-series, a distance correlation approach. J. Time Series Anal. 33, 438–457.
Zhou, Y., Xu, K., Zhu, L., Li, R., 2024. Rank-based indices for testing independence between two high-dimensional vectors. Ann. Statist. 52, 184–206.