Statistical Insights into HSIC in High Dimensions

Tao Zhang Yaowu Zhang Tingyou Zhou

1 Shanghai University of Finance and Economics
2 Zhejiang University of Finance and Economics

CONTENTS

1 Introduction

2 Asymptotic Properties in High Dimensions

3 Statistical Insights in High Dimensions

4 Numerical Studies

5 Conclusions

Introduction

Research Question

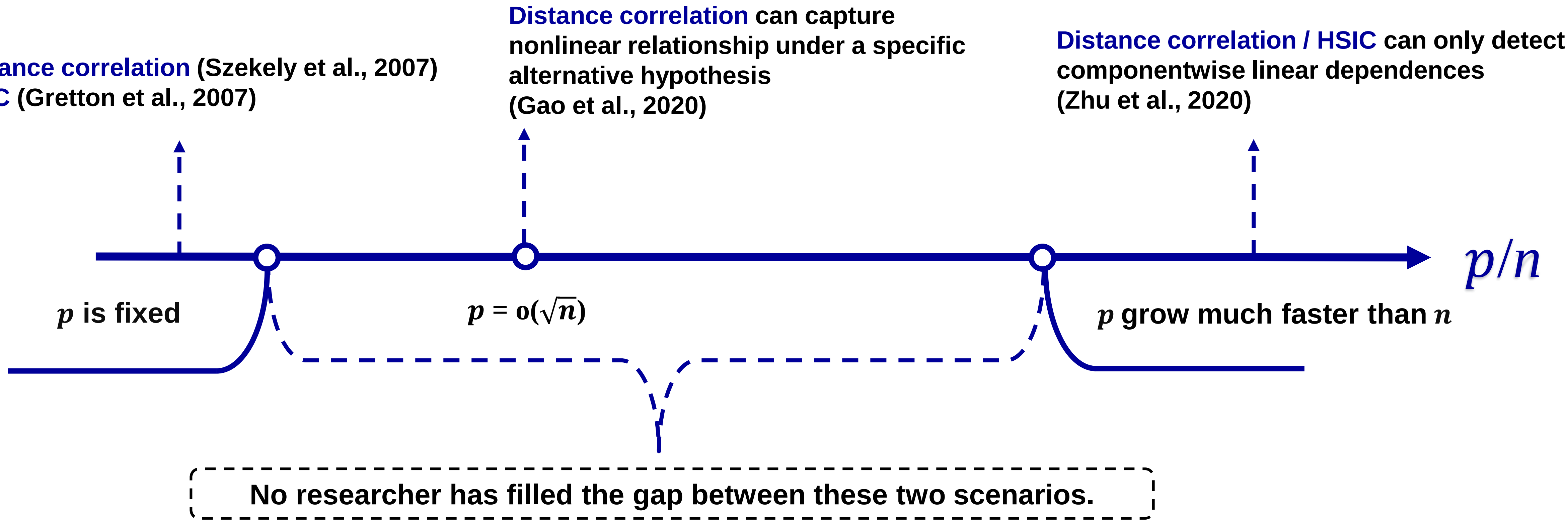
- Let $\mathbf{x} = \left(X_1, \dots, X_p\right)^T \in \mathbb{R}^p$ and $\mathbf{y} = \left(Y_1, \dots, Y_q\right)^T \in \mathbb{R}^q$ be two random vectors,

$$H_0: \mathbf{x} \text{ is independent of } \mathbf{y},$$
$$H_1: \mathbf{x} \text{ is not independent of } \mathbf{y}.$$

Value of Research

- The problems of measuring nonlinear dependence between $\mathbf{x}$ and $\mathbf{y}$ and testing for their independence are fundamental and have a wide range of applications in statistics and machine learning.

Related Literature



Motivation: So we want to bridge the gap between these two scenarios and provide statistical insights into the performance of HSIC when the dimensions grow at different rates.

Preliminaries

- The squared Hilbert-Schmidt norm

$$\mathrm{HSIC}(\mathbf{x},\mathbf{y}) = E\{K(\mathbf{x}_1,\mathbf{x}_2)L(\mathbf{y}_1,\mathbf{y}_2)\} + E\{K(\mathbf{x}_1,\mathbf{x}_2)\}E\{L(\mathbf{y}_1,\mathbf{y}_2)\} - 2E[\,E\{K(\mathbf{x}_1,\mathbf{x}_2) \mid \mathbf{x}_1\}E\{L(\mathbf{y}_1,\mathbf{y}_2) \mid \mathbf{y}_1\}\,].$$

- The squared sample Hilbert-Schmidt norm

$$\mathrm{HSIC}_n(\mathbf{x},\mathbf{y}) = \frac{1}{n(n-1)}\sum_{(i_1,i_2)} K(\mathbf{x}_{i_1},\ \mathbf{x}_{i_2})L(\mathbf{y}_{i_1},\ \mathbf{y}_{i_2}) - \frac{2}{n(n-1)(n-2)}\sum_{(i_1,i_2,i_3)} K(\mathbf{x}_{i_1},\ \mathbf{x}_{i_2})L(\mathbf{y}_{i_1},\ \mathbf{y}_{i_3})$$
$$+ \frac{1}{n(n-1)(n-2)(n-3)}\sum_{(i_1,i_2,i_3,i_4)} K(\mathbf{x}_{i_1},\ \mathbf{x}_{i_2})L(\mathbf{y}_{i_3},\ \mathbf{y}_{i_4}).$$

- The squared sample Hilbert-Schmidt correlation

$$\mathrm{hCorr}_n^2(\mathbf{x},\mathbf{y}) = \frac{\mathrm{HSIC}_n(\mathbf{x},\mathbf{y})}{\sqrt{\mathrm{HSIC}_n(\mathbf{x},\mathbf{x})\mathrm{HSIC}_n(\mathbf{y},\mathbf{y})}}.$$

Asymptotic properties in High Dimensions

- The asymptotic properties of the HSIC based test under **the null hypothesis**.

Theorem 1. *Assume the kernels are symmetric with finite fourth moment, i.e.,* $K(\mathbf{x}_1, \mathbf{x}_2) = K(\mathbf{x}_2, \mathbf{x}_1)$, $L(\mathbf{y}_1, \mathbf{y}_2) = L(\mathbf{y}_2, \mathbf{y}_1)$, $E\{K^4(\mathbf{x}_1, \mathbf{x}_2)\} < \infty$ *and* $E\{L^4(\mathbf{y}_1, \mathbf{y}_2)\} < \infty$. *Further assume that* $p + q \to \infty$,

$$\frac{E\{H_{\mathbf{x}}^4(\mathbf{x}_1, \mathbf{x}_2)\}E\{H_{\mathbf{y}}^4(\mathbf{y}_1, \mathbf{y}_2)\}}{n\{\mathrm{HSIC}(\mathbf{x}, \mathbf{x})\mathrm{HSIC}(\mathbf{y}, \mathbf{y})\}^2} \to 0, \quad \textit{and} \quad \frac{E\{G_{\mathbf{x}}^2(\mathbf{x}_1, \mathbf{x}_2)\}E\{G_{\mathbf{y}}^2(\mathbf{y}_1, \mathbf{y}_2)\}}{\{\mathrm{HSIC}(\mathbf{x}, \mathbf{x})\mathrm{HSIC}(\mathbf{y}, \mathbf{y})\}^2} \to 0,$$

as $n \to \infty$. *Then under the null hypothesis, we have* $2^{-1/2} n \,\mathrm{hCorr}_n^2(\mathbf{x}, \mathbf{y}) \xrightarrow{d} N(0, 1)$.

- It greatly expedites the implementation of HSIC based tests because no additional permutations are required to decide critical values.

Asymptotic properties in High Dimensions

- The two assumptions of Theorem 2.

(A1) There exists some $\kappa_{\mathbf{z}} > 0$ such that $E\{\|\mathbf{z}^*\|^2 - E(\|\mathbf{z}^*\|^2)\}^{2k} \asymp E(\mathbf{z}_1^{*\mathrm{T}}\mathbf{z}_2^*)^{2k} \asymp d^{-k\kappa_{\mathbf{z}}}$ for all $k \in \mathbb{N}^+$.

(A2) Let $k_0(x) = k(x^{1/2})$ and $l_0(y) = l(y^{1/2})$. The first and second derivatives of $k_0(\cdot)$ and $l_0(\cdot)$ are uniformly bounded away from zero to infinity around $E\|\mathbf{x}_1^* - \mathbf{x}_2^*\|^2$ and $E\|\mathbf{y}_1^* - \mathbf{y}_2^*\|^2$, respectively.

- The power performance of the HSIC based test under **<u>the alternative hypothesis</u>**.

Theorem 2. *Assume (A1) and (A2) hold true. Then under the alternative hypothesis, if* $n^{1/2}\mathrm{hCorr}^2(\mathbf{x}, \mathbf{y}) \to \infty$ *as* $n \to \infty$, *we have* $n\,\mathrm{hCorr}_n^2(\mathbf{x}, \mathbf{y}) \to \infty$ *in probability.*

- Theorem 2 guarantee that the HSIC based test can have nontrivial power in high dimensions together as long as the signal strength does not decay to zero too fast.

Statistical Insights in High Dimensions

- We expand $\mathrm{HSIC}(\mathbf{x},\mathbf{y})$ at the population level

Theorem 3. *Assume (A1) and (A2) hold true. Then under the alternative hypothesis,*

1. *when $p\to\infty$ and q remains fixed as $n\to\infty$, if $E(\mathbf{x}^{\otimes t}\mid \mathbf{y}) = E(\mathbf{x}^{\otimes t})$ hold true for all $t<s$ for some $s\in\mathbb{N}^+$, then $\mathrm{HSIC}(\mathbf{x},\mathbf{y}) = O(p^{-s\kappa_{\mathbf{x}}/2})$, and*

$$\mathrm{HSIC}(\mathbf{x},\mathbf{y}) \quad = \quad k_0^{(s)} \sum_{2a+c=s} \frac{(-2)^c}{a!a!c!}\mathrm{MD}^2(\mathbf{x}^{*\otimes c}\|\mathbf{x}^*\|^{2a}\mid \mathbf{y}) + o(p^{-s\kappa_{\mathbf{x}}/2}),$$

2. *when $p\to\infty$ and $q\to\infty$ as $n\to\infty$, if $\mathrm{cov}(\mathbf{x}^{\otimes t_1},\mathbf{y}^{\otimes t_2})\neq \mathbf{0}$ only when $t_1\geq s_1$ and $t_2\geq s_2$ for some $s_1,s_2\in\mathbb{N}^+$, then $\mathrm{HSIC}(\mathbf{x},\mathbf{y}) = O(p^{-s_1\kappa_{\mathbf{x}}/2}q^{-s_2\kappa_{\mathbf{y}}/2})$, and*

$$\begin{aligned}\mathrm{HSIC}(\mathbf{x},\mathbf{y}) = &\sum_{2a_1+c_1=s_1}\sum_{2a_2+c_2=s_2} \frac{k_0^{(s_1)}(-2)^{c_1}}{a_1!a_1!c_1!}\frac{l_0^{(s_2)}(-2)^{c_2}}{a_2!a_2!c_2!}\left\|\mathrm{cov}\{\|\mathbf{x}^*\|^{2a_1}\mathbf{x}^{*\otimes c_1},\|\mathbf{y}^*\|^{2a_2}\mathbf{y}^{*\otimes c_2 T}\}\right\|_F^2\\ &+o(p^{-s_1\kappa_{\mathbf{x}}/2}q^{-s_2\kappa_{\mathbf{y}}/2}).\end{aligned}$$

- The Theorem 3 characterizes the changing process of the ability to measure nonlinear dependence in high dimensions for HSIC.

Statistical Insights in High Dimensions

Theorem 2 and Proposition 1

$$n^{1/2}p^{\kappa_x/2}q^{\kappa_y/2}\mathrm{HSIC}(\mathbf{x},\mathbf{y}) \to \infty$$

HSIC based test have asymptotic power 1

Theorem 3

The association type between random vectors

The dependence structures within covariates

$$\boldsymbol{p^{(s-1)\kappa_x} = o(n)}$$

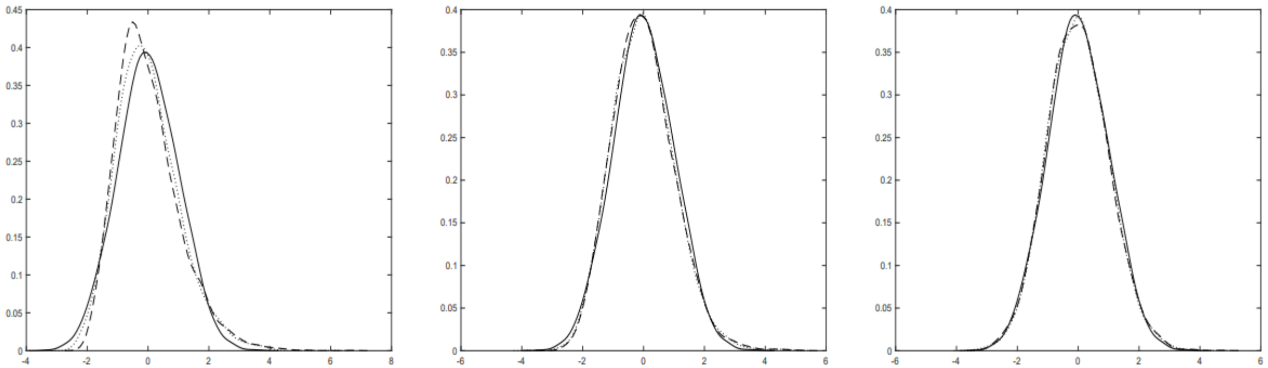
Sample size

Dimensionality

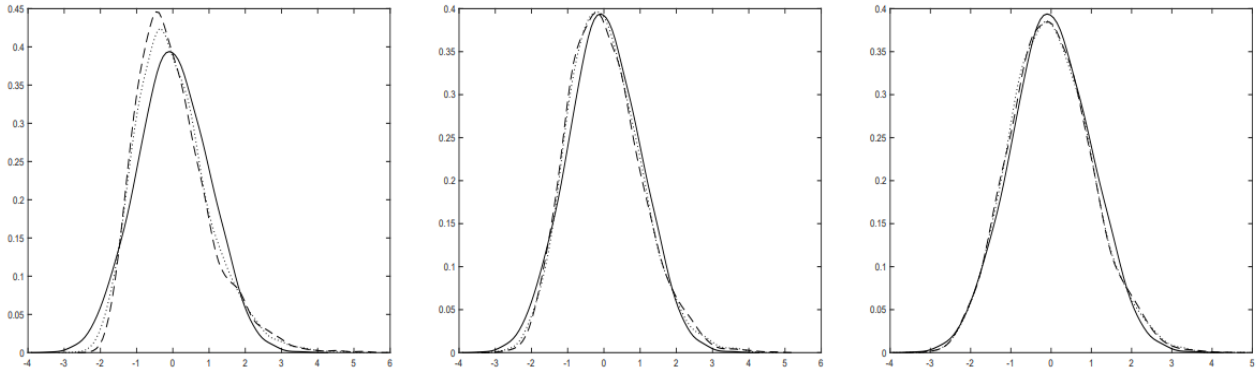
$$\boldsymbol{p^{(s_1-1)\kappa_x}q^{(s_2-1)\kappa_y} = o(n)}$$

Numerical Studies

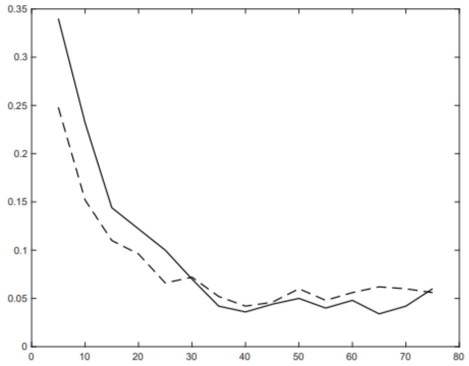
Example 1. Let n = 100, $\Sigma_{\mathbf{x}} = \mathbf{I}_{p\times p}$ and $\Sigma_{\mathbf{y}} = \mathbf{I}_{q\times q}$, we generate $\mathbf{x} = \left(X_1, \dots, X_p\right)^T \in \mathbb{R}^p$and $\mathbf{y} = \left(Y_1, \dots, Y_q\right)^T \in \mathbb{R}^q$ and $\mathbf{x}\sim$ N(0, $\Sigma_{\mathbf{x}}$) and $\mathbf{y}\sim$ N(0, $\Sigma_{\mathbf{y}}$) be independent. We consider two scenario : (1) q =1 and vary p from {5, 25, 100}; (2) p = q = d and vary d from {2, 5, 10}.



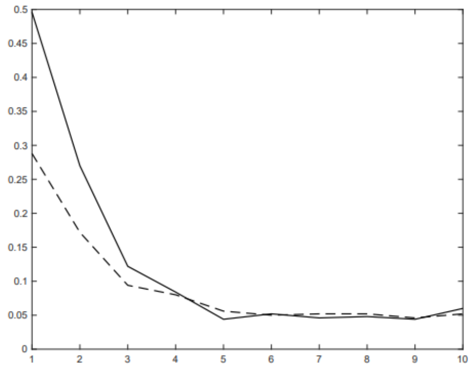
(A): $p = 5$ (B): $p = 25$ (C): $p = 100$



(A): $d = 2$ (B): $d = 5$ (C): $d = 10$



(A): The first scenario.



(B): The second scenario.

Numerical Studies

Example 2. Let n =100, $\Sigma_{\mathbf{x}} = \left(0.5^{|i-j|}\right)_{p\times p}$, we generate $\mathbf{x} = \left(X_1, \ldots, X_p\right)^T \in \mathbb{R}^p$ and $\mathbf{x} \sim$ N(0, $\Sigma_{\mathbf{x}}$), fix q = 1 and vary p from {30, 50, 100, 200, 500, 1000}, The independent error term ε follows standard normal distribution and the univariate response Y is generated through

Model (I): $Y = X_1 + \cdots + X_p + \varepsilon;$

Model (II): $Y = X_1^2 + \cdots + X_p^2 + \varepsilon;$

Model (III): $\{(X_{2k-1}, X_{2k})^T \mid Y\} \sim N\left\{\mathbf{0}, \left(\rho_{k,Y}^{|i-j|}\right)_{2\times 2}\right\}, k = 1, \ldots, p/2.$

Model	Test	p					
		30	50	100	200	500	1000
(I)	Gaussian	1.000	1.000	1.000	1.000	0.998	0.954
	Laplacian	1.000	1.000	1.000	1.000	0.994	0.916
	DC	1.000	1.000	1.000	1.000	1.000	0.998
(II)	Gaussian	0.934	0.774	0.484	0.318	0.184	0.132
	Laplacian	0.998	0.978	0.834	0.596	0.314	0.234
	DC	0.974	0.894	0.650	0.412	0.226	0.176
(III)	Gaussian	0.044	0.050	0.060	0.046	0.060	0.068
	Laplacian	0.050	0.046	0.054	0.048	0.054	0.060
	DC	0.050	0.050	0.052	0.034	0.064	0.044

Numerical Studies

Example 3. Let n =100, $\Sigma_{\mathbf{x}} = \left(0.5^{|i-j|}\right)_{p\times p}$, we generate $\mathbf{x} = \left(X_1, \ldots, X_p\right)^T \in \mathbb{R}^p$ and $\mathbf{x} \sim$ N(0, $\Sigma_{\mathbf{x}}$). We set p = q = d and vary d from {6, 10, 20, 50, 100, 200}. The independent error terms $\varepsilon_1, \ldots, \varepsilon_d$ are generated from d independent standard normal distributions and the $\mathbf{y} = (Y_1, \ldots, Y_d)^T$ is generated through

Model (IV): $Y_j = X_j + \varepsilon_j, j = 1, \ldots, d;$

Model (V): $Y_j = X_j^2 + \varepsilon_j, j = 1, \ldots, d;$

Model (VI): $\{(X_{2k-1}, X_{2k})^T \mid \mathbf{y}\} \sim N\left\{\mathbf{0}, \left(\rho_{k,\mathbf{y}}^{|i-j|}\right)_{2\times 2}\right\}, k = 1, \ldots, d/2.$

Model	Test	d					
		6	10	20	50	100	200
(IV)	Gaussian	1.000	1.000	1.000	1.000	1.000	1.000
	Laplacian	1.000	1.000	1.000	1.000	1.000	1.000
	DC	1.000	1.000	1.000	1.000	1.000	1.000
(V)	Gaussian	1.000	1.000	0.944	0.440	0.242	0.140
	Laplacian	1.000	1.000	1.000	0.904	0.578	0.282
	DC	1.000	0.986	0.844	0.400	0.232	0.136
(VI)	Gaussian	0.056	0.064	0.046	0.078	0.038	0.072
	Laplacian	0.058	0.064	0.044	0.074	0.040	0.072
	DC	0.060	0.066	0.046	0.078	0.040	0.072

Real Data Applications

- There exists dependences between the monthly mean stock prices of the energy sector and the raw material sector from the results.

$\mathbf{x}$: Stock returns series of 224 companies from the raw material sector.
$\mathbf{y}$: Stock returns series of 214 companies from the energy sector.

Gaussian kernels p-values: 2.031×10^{-10}
Laplacian kernels p-values: 2.749×10^{-9}
RV coefficient p-values: 2.02×10^{-4}

- The average stock returns for other software companies may change depending on how the leading software companies perform.

$\mathbf{x}$: Stock return series of Mercado Libre and Microsoft.
$\mathbf{y}$: Stock returns series of 259 software & service companies.

Gaussian kernels p-values: 7.438×10^{-5}
Laplacian kernels p-values: 4.954×10^{-6}
RV coefficient p-values: 0.0584

Conclusions

- The asymptotic null distribution of a rescaled HSIC is a **standard normal** in the high dimensional setting.

- The general condition for the HSIC based tests to have **power asymptotically approaching one**.

- This condition depends on **the sample size, the covariate dimensions, the dependence structures within covariates, and the association types** between $\mathbf{x}$ and $\mathbf{y}$.

THANK YOU