Conditional Diffusion Models Based Conditional Independence Testing

Yanfeng Yang[1*], Shuai Li[1*], Yingjie Zhang[1*], Zhuoran Sun[1], Hai Shu[2], Ziqi Chen[1†], Renming Zhang[3]

[1]School of Statistics, KLATASDS-MOE, East China Normal University, Shanghai, China
[2] Department of Biostatistics, School of Global Public Health, New York University, New York, USA
[3] Department of Computer Science, Boston University, Boston, USA

Abstract

Conditional independence (CI) testing is a fundamental task in modern statistics and machine learning. The conditional randomization test (CRT) was recently introduced to test whether two random variables, X and Y, are conditionally independent given a potentially high-dimensional set of random variables, Z. The CRT operates exceptionally well under the assumption that the conditional distribution $X|Z$ is known. However, since this distribution is typically unknown in practice, accurately approximating it becomes crucial. In this paper, we propose using conditional diffusion models (CDMs) to learn the distribution of $X|Z$. Theoretically and empirically, it is shown that CDMs closely approximate the true conditional distribution. Furthermore, CDMs offer a more accurate approximation of $X|Z$ compared to GANs, potentially leading to a CRT that performs better than those based on GANs. To accommodate complex dependency structures, we utilize a computationally efficient classifier-based conditional mutual information (CMI) estimator as our test statistic. The proposed testing procedure performs effectively without requiring assumptions about specific distribution forms or feature dependencies, and is capable of handling mixed-type conditioning sets that include both continuous and discrete variables. Theoretical analysis shows that our proposed test achieves a valid control of the type I error. A series of experiments on synthetic data demonstrates that our new test effectively controls both type-I and type-II errors, even in high dimensional scenarios.

Code — https://github.com/Yanfeng-Yang-0316/CDCIT

Introduction

Conditional independence (CI) is an important concept in statistics and machine learning. Testing conditional independence plays a central role in classical problems such as causal inference (Pearl 1988; Spirtes, Glymour, and Scheines 2000), graphical models (Lauritzen 1996; Koller and Friedman 2009), and variable selection (Dai, Shen, and Pan 2022). It is widely used in various scientific problems, including gene regulatory network inference (Dai et al. 2024) and personalized therapies (Khera and Kathiresan 2017; Zhu et al. 2018). We consider testing whether two random variables, X and Y, are independent given a random vector Z, based on observations of the joint density $p_{X,Y,Z}(x, y, z)$. Specifically, we test the following hypotheses:

$$H_0 : X \perp\!\!\!\perp Y|Z \text{ versus } H_1 : X \not\perp\!\!\!\perp Y|Z,$$

where $\perp\!\!\!\perp$ denotes the independence. In practical genome-wide association studies, X represents a specific genetic variant, Y denotes disease status, and Z accounts for the rest of the genome. By conditioning on Z, we can evaluate whether the genetic variant X has an effect on the disease status Y (Liu et al. 2022) by CI testing. CI testing becomes particularly challenging due to the high-dimensionality of the conditioning vector Z (Bellot and van der Schaar 2019; Shi et al. 2021). Moerever, the presence of mixed discrete and continuous variables in Z as in many real-world applications presents further challenges for the testing (Mesner and Shalizi 2020; Zan et al. 2022).

Recently, there has been a large and growing literature on CI testing, and for a more comprehensive review, we refer readers to Li and Fan (2020). The metric-based tests (e.g., Su and White (2008), Su and White (2014), Wang et al. (2015)) employ some kernel smoothers to estimate the conditional characteristic function or the distribution function of Y given X and Z. However, due to the curse of dimensionality, these tests are usually not suitable when the conditioning vector Z is high-dimensional. The kernel-based tests, such as Fukumizu et al. (2007), Zhang et al. (2011), and Scetbon, Meunier, and Romano (2022), represent probability distributions as elements of a reproducing kernel Hilbert space (RKHS), which enables us to understand properties of these distributions using Hilbert space operations. However, these tests based on asymptotic distributions may exhibit inflated type-I errors or inadequate power when dealing with high-dimensional Z (Doran et al. 2014; Runge 2018; Shi et al. 2021). The most relevant work to ours is the conditional randomization test (CRT) proposed by Candès et al. (2018), which assumes the true conditional distribution of X given Z, denoted by $P(\cdot|Z)$, is known. It is theoretically proven that the CRT maintains validity by ensuring that the type I error does not exceed the significance level α (Berrett et al. 2020; Liu et al. 2022).

However, the true conditional distribution $P(\cdot|Z)$ is rarely

*These authors contributed equally.
†Corresponding author: Ziqi Chen (zqchen@fem.ecnu.edu.cn, chenzq453@163.com).
Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

known in practice, several methods have been developed to approximate the $P(\cdot|Z)$. The smoothing-based methods (Hall, Racine, and Li 2004; Hall and Yao 2005; Izbicki and Lee 2017) suffer from the curse of dimensionality, and their performance deteriorates sharply when the dimensionality of Z becomes large. Bellot and van der Schaar (2019) developed a Generative Conditional Independence Test (GCIT) by using Wasserstein generative adversarial networks (WGANs; Arjovsky, Chintala, and Bottou 2017) to approximate $P(\cdot|Z)$. Shi et al. (2021) proposed to use the Sinkhorn GANs (Genevay, Peyré, and Cuturi 2018) to approximate $P(\cdot|Z)$. However, the training of GANs is often unstable, with the risk of collapse if hyperparameters and regularizers are not carefully chosen (Dhariwal and Nichol 2021). The potentially inaccurate learning of conditional distributions using GANs can lead to inflated type I errors in GANs-based CI tests (Li et al. 2023). Li et al. (2023) introduced a method using the 1-nearest neighbor technique to generate samples from the approximated conditional distribution of X given Z. However, their approach necessitates dividing the dataset into two segments. Consequently, only one-third of the total samples are allocated to the testing dataset used for calculating test statistics, which reduces the test's statistical power. Li et al. (2024) proposed utilizing a k-nearest-neighbor local sampling strategy to generate samples from the approximated conditional distribution of X given Z. Nevertheless, this approach encounters issues with insufficient sample diversity, resulting in unstable performance of the CI test, particularly when the conditioning variables Z include both continuous and discrete variables.

Diffusion models, which have recently emerged as a notable class of generative models, have attracted significant attention (Ho, Jain, and Abbeel 2020; Song et al. 2021; Yang et al. 2023). Unlike GANs, diffusion models provide a much more stable training process and generate more realistic samples (Dhariwal and Nichol 2021; Song, Meng, and Ermon 2021). They have achieved great success in various tasks, such as image generation (Ho, Jain, and Abbeel 2020) and video generation (Ho et al. 2022). In this paper, we propose using conditional diffusion models to approximate $P(\cdot|Z)$ and generating samples from the approximated conditional distribution. Moreover, as highlighted in Li et al. (2023), the choice of test statistics in CRT procedure is crucial for achieving adequate statistical power as well as controlling type I errors. Conditional mutual information (CMI) for (X, Y, Z), denoted as $I(X; Y|Z)$, provides a strong theoretical guarantee for conditional dependence relations. Specifically, $I(X; Y|Z) = 0 \iff X \perp\!\!\!\perp Y|Z$ (Cover and Thomas 2012). In this paper, we adopt the classifier-based CMI estimator as the test statistic (Mukherjee, Asnani, and Kannan 2020; Li et al. 2024).

Our main contributions are summarized as follows. First, we propose, for the first time, using conditional diffusion models to generate samples for the conditional distribution in the CI testing task. Compared to GANs, this method is not only more stable during training but also demonstrates significant advantages in sample quality and diversity. It is theoretically and empirically shown that the distribution of the generated samples is very close to the true conditional distribution. Second, we use a computationally efficient classifier-based CMI estimator as the test statistic, which captures intricate dependence structures among variables. Third, theoretical analysis demonstrates that our proposed test achieves a valid control of the type I error asymptotically. Fourth, our empirical evidence demonstrates that, compared to state-of-the-art methods, our test effectively controls type I error while maintaining sufficient power under H_1. This remains true even when handling high-dimensional data and/or mixed-type conditioning sets that include both continuous and discrete variables.

The Proposed Approach

The Conditional Randomization Test (CRT)

Our work builds on the conditional randomization test (CRT) proposed by Candès et al. (2018), which, however, assumes that the true conditional distribution $P(\cdot|Z)$ is known. Specifically, consider n i.i.d. copies $\mathcal{D}_{\mathrm{T}} = \{(X_i, Y_i, Z_i) : i = 1, \ldots, n\}$ of (X, Y, Z). If $P(\cdot|Z)$ is known, conditionally on $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^T$, one can independently draw $X_i^{(b)} \sim P(\cdot|Z_i)$ for each i across $b = 1, \ldots, B$ such that all $\boldsymbol{X}^{(b)} := (X_1^{(b)}, \ldots, X_n^{(b)})^T$ are independent of $\boldsymbol{X} := (X_1, \ldots, X_n)^T$ and $\boldsymbol{Y} := (Y_1, \ldots, Y_n)^T$, where B is the number of repetitions. Thus, under the null hypothesis $H_0 : X \perp\!\!\!\perp Y|Z$, we have $(\boldsymbol{X}^{(b)}, \boldsymbol{Y}, \boldsymbol{Z}) \overset{d}{=} (\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ for all b, where $\overset{d}{=}$ denotes equality in distribution. A large difference between $(\boldsymbol{X}^{(b)}, \boldsymbol{Y}, \boldsymbol{Z})$ and $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ can be regarded as a strong evidence against H_0. Statistically, for any test statistic $\boldsymbol{T}(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$, we can calculate the p-value of the CI test by

$$\frac{1 + \sum_{b=1}^{B} \mathbf{I}\{\boldsymbol{T}(\boldsymbol{X}^{(b)}, \boldsymbol{Y}, \boldsymbol{Z}) \geq \boldsymbol{T}(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})\}}{1 + B}, \quad (1)$$

where $\mathbf{I}\{\cdot\}$ is the indicator function. Under the null hypothesis H_0, since the $(B + 1)$ triples $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}), (\boldsymbol{X}^{(1)}, \boldsymbol{Y}, \boldsymbol{Z}), \ldots, (\boldsymbol{X}^{(B)}, \boldsymbol{Y}, \boldsymbol{Z})$ are exchangeable, the above p-value is valid. Specifically, $P(p \leq \alpha|H_0) \leq \alpha$ holds for any $\alpha \in (0, 1)$ (Candès et al. 2018; Berrett et al. 2020).

Methodology for Sampling

The traditional CRT procedure assumes that the true conditional distribution $P(\cdot|Z)$ is known. However, in practice, this distribution is seldomly known. We propose using score-based conditional diffusion models to approximate the conditional distribution of X given Z. Specifically, we have an unlabelled data set $\mathcal{D}_{\mathrm{U}}$ that consists of N i.i.d. samples $\{(X_i^{\mathrm{U}}, Z_i^{\mathrm{U}}) : i = 1, \ldots, N\}$ from the distribution $P_{X,Z}$. We aim to accurately recover the true distribution of X conditioning on Z using $\mathcal{D}_{\mathrm{U}}$.

The score-based conditional diffusion models involve two stochastic processes: the forward process and the reverse process. Specifically, let $t \in [0, T]$ be the time index in forward process and reverse process, and denote $X := X(0) \in \mathbb{R}^{d_x}$. Conditioned on Z, the forward process is presented as:

$$dX(t) = -(X(t)/2)dt + dB(t),$$

where $X(0) \sim P(\cdot|Z)$ and $B(t)$ is a standard Brownian motion in $\mathbb{R}^{d_x}$. At any time t, let $p_t(\cdot|Z)$ and $P_t(\cdot|Z)$ be the conditional density and distribution of $X(t)|Z$, respectively. By the property of Ornstein Uhlenbeck (OU) process, we derive

$$X(t) \stackrel{d}{=} \exp(-t/2)X(0) + \sqrt{1-\exp(-t)}\epsilon, \qquad (2)$$

where $\epsilon \sim N(0, I_{d_x})$ and I_{d_x} is the d_x-dimensional identity matrix. It can be deduced that as T approaches infinity, $X(T) \sim N(0, I_{d_x})$. In practice, however, $X(t)$ is stopped at a sufficiently large T to ensure computational feasibility.

According to Song et al. (2021), the true reverse process is:

$$d\overline{X}(t) = \left[\frac{1}{2}\overline{X}(t) + \nabla \log p_{T-t}(\overline{X}(t)|Z)\right]dt + d\overleftarrow{B}(t),$$
$$\overline{X}(0) \sim P_T(\cdot|Z),$$

where $\overleftarrow{B}(t)$ is a standard Brownian motion in reversed process. We observe that the distributions of $\overline{X}(t)$ and $X(T-t)$ are identical. Thus, the conditional density and distribution of $\overline{X}(t)|Z$ are $p_{T-t}(\cdot|Z)$ and $P_{T-t}(\cdot|Z)$, respectively. $\nabla \log p_{T-t}(\cdot|z)$ is the score function of $\overline{X}(t)$ conditioned on $Z = z$. The analytical solution for the conditional score function is unattainable. Here, we utilize a ReLU neural network $\widehat{s}(x, z, t) \in \mathcal{F}$ to approximate it, where $\mathcal{F}$ is the function class of ReLU neural networks. Following Ho, Jain, and Abbeel (2020) and Song et al. (2021), we train $\widehat{s}(x, z, t)$ by approximating the conditional score function through minimization of the following objective:

$$\mathbb{E}_t \mathbb{E}_{X(0),Z} \mathbb{E}_{X(t)} \|\widehat{s}(X(t), Z, t) - \nabla \log \phi(X(t), X(0))\|_2^2,$$

where

$$X(t) \sim N(\exp(-t/2)X(0),\ [1-\exp(-t)]I_{d_x}),$$
$$\phi(X(t), X(0)) = \exp\left[\frac{\|X(t) - \exp(-t/2)X(0)\|^2}{-2(1-\exp(-t))}\right],$$
$$(X(0), Z) \sim P_{X,Z},\ t \sim \text{Uniform}[t_{\min}, T],$$

and $t_{\min}$ is an early-stopping time close to zero ensuring the denominator in ϕ is not zero. As described in (2), when T is sufficiently large, $P_T(\cdot|Z)$ can be approximated by $N(0, I_{d_x})$. Therefore, given Z, we propose the following reverse process with the approximated score function to generate pseudo samples:

$$d\overleftarrow{X}(t) = \left[\frac{1}{2}\overleftarrow{X}(t) + \widehat{s}(\overleftarrow{X}(t), Z, T-t)\right]dt + d\overleftarrow{B}(t),$$
$$\overleftarrow{X}(0) \sim N(0, I_{d_x}). \qquad (3)$$

Let $\overleftarrow{X}(T - t_{\min})$ be the pseudo-sample generated by (3). As demonstrated in Theorem 1, this pseudo-sample has a conditional distribution $\widehat{P}(\cdot|Z)$ that approximates the true conditional distribution $P(\cdot|Z)$ effectively (Fu et al. 2024). Algorithm 1 outlines the training procedure for the conditional score matching models, whereas the sampling process from the reverse process is detailed in Algorithm 2.

Algorithm 1: Training the conditional score matching models

Input: A data set with N i.i.d. samples $\{X_i^{\text{U}}, Z_i^{\text{U}}\}_{i=1}^N$.
Output: The score network $\widehat{s}$.

Let $\{X_i(0)\}_{i=1}^N = \{X_i^{\text{U}}\}_{i=1}^N$
Initialize a deep neural network $\widehat{s}(x, z, t)$
while not converge **do**
  Draw $t \sim \text{Uniform}[t_{\min}, T]$
  Draw $\epsilon_1, \ldots, \epsilon_N \sim N(0, I_{d_x})$
  Let $X_i(t) = \exp(-t/2)X_i(0) + \sqrt{1-\exp(-t)}\epsilon_i$
  Compute $L_{\text{score}} = \sum_{i=1}^N \|\widehat{s}(X_i(t), Z_i^{\text{U}}, t) + [X_i(t) - \exp(-t/2)X_i(0)]/(1-\exp(-t))\|_2^2$
  Take optimization step on ∇L_{score} and update the parameters of $\widehat{s}$
end while
Return $\widehat{s}$

Algorithm 2: Sampling from score-based conditional diffusion models

Input: A sample $Z \sim P_Z$, the score network $\widehat{s}$, and sample step K.
Output: Pseudo sample $\widehat{X}$.

Evenly divide $[0, T - t_{\min}]$ into $t_0 = 0 < t_1 < \ldots < t_K = T - t_{\min}$ and let $\Delta t = (T - t_{\min})/K$
Draw $\overleftarrow{X}(0) \sim N(0, I_{d_x})$.
for $k = 0$ to $K - 1$ **do**
  Draw $\epsilon_k \sim N(0, I_{d_x})$
  Let $\overleftarrow{X}(k+1) = \overleftarrow{X}(k) + \sqrt{\Delta t}\epsilon_k + \left[\frac{1}{2}\overleftarrow{X}(k) + \widehat{s}(\overleftarrow{X}(k), Z, T - t_k)\right]\Delta t$
end for
Let $\widehat{X} = \overleftarrow{X}(K)$
Return $\widehat{X}$

Theoretical Guarantee for Sampling Quality

We present a theoretical result showing that the distribution of the generated samples closely resembles the true conditional distribution. Specifically, denote the true conditional density of X given Z as $p(\cdot|Z)$, and the density of $\widehat{X}$ sampled from Algorithm 2 given Z as $\widehat{p}(\cdot|Z)$. The total variation distance between two density functions p_1 and p_2 is defined as $d_{\text{TV}}(p_1, p_2) = \frac{1}{2}\int |p_1(x) - p_2(x)|dx$. Define

$$\Gamma_1(k, \alpha) = \frac{k+\alpha}{d_x + d_z + 2k + 2\alpha},$$
$$\Gamma_2(k, \alpha) = \max\left(\frac{19}{2}, \frac{k+\alpha+2}{2}\right),$$

where k and α are defined in Supplementary Materials. The proof of Theorem 1 is deferred to Supplementary Materials.

Theorem 1. *Under Assumptions 1 and 2 in the Supplementary Material, taking early stopping time* $t_{\min} = N^{-4\Gamma_1(k,\alpha)-1}$ *and terminal time* $T = 2\Gamma_1(k, \alpha)\log N$, *when* $N \to \infty$, *we have*

$$\mathbb{E}_{\{X_i^{\mathrm{U}}, Z_i^{\mathrm{U}}\}_{i=1}^N} \mathbb{E}_Z[d_{\mathrm{TV}}(p(\cdot|Z), \widehat{p}(\cdot|Z))]$$
$$= O\left(N^{-\Gamma_1(k,\alpha)} \cdot (\log N)^{\Gamma_2(k,\alpha)}\right).$$

Empirical Evidence for Sampling Quality

In this subsection, we investigate the empirical performance of our proposed method for approximating the conditional distribution in sample generation in the CRT framework. Specifically, we compare it with several well-known existing conditional distribution estimating methods in CI test, including WGANs (Bellot and van der Schaar 2019), Sinkhorn GANs (Shi et al. 2021), and k-nearest neighbors (k-NN) (Li et al. 2024). We consider the following three models.

- Model M1: $X = Z_1^2 + \exp(Z_2 + Z_3/3) + Z_4 - Z_5 + 0.5(1 + Z_2^2 + Z_5^2)\epsilon$, where $Z_1, \ldots, Z_5, \epsilon \overset{i.i.d.}{\sim} N(0,1)$.
- Model M2: $X = (5 + Z_1^2/3 + Z_2^2 + Z_3^2 + Z_4^2 + Z_5^2)\exp(r)$, where $r \sim B \times N(-2,1) + (1-B) \times N(2,1)$, $B \sim$ Bernoulli$(1, 0.5)$, and $Z_1, \ldots, Z_5 \overset{i.i.d.}{\sim} N(0,1)$.
- Model M3: $X = \sum_{i=1}^{13} Z_i/13 + 0.33\epsilon$, where $Z_1, \ldots, Z_{10}, \epsilon \overset{i.i.d.}{\sim} N(0,1)$, and $Z_{11}, \ldots, Z_{20} \overset{i.i.d.}{\sim} 2 \cdot$ Bernoulli$(1, 0.5) - 1$.

Models M1 and M2 exhibit complex, nonlinear and non-monotonic relationships between X and Z. Model M3 involves a mixed-type Z consisting of both continuous and discrete variables, where we model the true relationship between X and $Z = (Z_1, \ldots, Z_{20})$ using only $(Z_1, \ldots, Z_{13})$. For each model, we use 500 samples to learn the conditional distribution for each method.

For each model, we estimate the conditional density functions using the 500 samples generated from each method with kernel smoothing (Weglarczyk 2018). Figure 1 and Figure 3 (Supplementary Material) display the estimated conditional density functions for a randomly generated value of Z. The results demonstrate that our conditional diffusion estimation method yields better conditional density estimators than WGANs, Sinkhorn GANs and k-NN. Further, it can be observed that the samples generated by k-NN lack diversity.

To further evaluate these methods, we compute the mean squared errors between the quantiles of the generated samples and those of the true conditional distribution. Specifically, a value z is randomly sampled. Given $Z = z$, we generate 500 samples for each conditional distribution estimating method. We then calculate the squared error between the τ quantile of the generated samples and the τ quantile of the true distribution over $\tau \in \{0.05, 0.25, 0.50, 0.75, 0.95\}$. This procedure is repeated 100 times, and the mean squared errors for each τ are reported in Table 1. Our proposed sampling method performs best on Models M1 and M3. It is also quite competitive on Model M2, though it is not always the top performer.

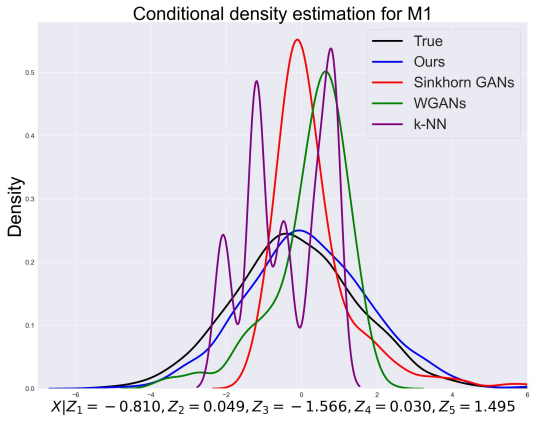

Figure 1: Comparison of conditional density estimators on Model M1. $Z = (-0.810, 0.049, -1.566, 0.030, 1.495)$.

Test Statistic

Conditional mutual information (CMI) $I(X;Y|Z)$ is defined as

$$\iiint p_{X,Y,Z}(x,y,z) \log \frac{p_{X,Y,Z}(x,y,z)}{p_{X,Z}(x,z)p_{Y|Z}(y|z)} dxdydz,$$

where $p_{X,Y,Z}(x,y,z)$ is the joint density of (X,Y,Z), $p_{X,Z}(x,z)$ is the joint density of (X,Z), and $p_{Y|Z}(y|z)$ is the conditional density of Y given $Z = z$. CMI as a tool for measuring conditional dependency, with $I(X;Y|Z) = 0 \iff X \perp\!\!\!\perp Y|Z$, has been used in conditional independence testing (Runge 2018; Li et al. 2024). Its major advantages include not needing the data to adhere to particular distributional assumptions or the features to have specific dependency relationships, making CMI applicable to a wide range of real-world datasets (Mukherjee, Asnani, and Kannan 2020).

In Equation (1), our test statistic $\boldsymbol{T}(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ is set as the classifier-based CMI estimator (CCMI, Mukherjee, Asnani, and Kannan, 2020). Specifically, $I(X;Y|Z)$ can be expressed in terms of Kullback-Leibler (KL) divergence:

$$I(X;Y|Z) = d_{\mathrm{KL}}(p_{X,Y,Z}(x,y,z)||p_{X,Z}(x,z)p_{Y|Z}(y|z)), \tag{4}$$

where $d_{\mathrm{KL}}(f||g)$ denotes the KL divergence between two distribution functions F and G, with density functions $f(x)$ and $g(x)$, respectively. We further utilize the Donsker-Varadhan (DV) representation of $d_{\mathrm{KL}}(f||g)$,

$$\sup_{s \in \mathcal{S}} \left[\mathbb{E}_{w \sim f} s(w) - \log \mathbb{E}_{w \sim g} \exp\{s(w)\}\right], \tag{5}$$

where the function class $\mathcal{S}$ includes all functions with finite expectations. In fact, the optimal function in (5) is given by $s^*(x) = \log\{f(x)/g(x)\}$ (Belghazi et al. 2018), which leads to:

$$d_{\mathrm{KL}}(f||g) = \mathbb{E}_{w \sim f} \log\{f(w)/g(w)\} - \log[\mathbb{E}_{w \sim g}\{f(w)/g(w)\}]. \tag{6}$$

Next, our primary goal is to empirically estimate (6) with $f = p_{X,Y,Z}(x,y,z)$ and $g = p_{X,Z}(x,z)p_{Y|Z}(y|z)$,

Model	Quantile	Ours	WGANs	Sinkhorn GANs	k-NN
M1	0.05	**1.731(3.740)**	7.478(6.372)	5.474(5.910)	2.062(5.962)
	0.25	**0.508(0.998)**	3.676(4.585)	2.384(2.468)	1.201(2.193)
	0.50	**0.405(0.421)**	2.568(3.408)	2.595(2.905)	0.876(1.773)
	0.75	**0.951(0.936)**	2.364(2.509)	4.451(4.221)	1.601(2.278)
	0.95	**3.758**(4.589)	3.871(**1.386**)	9.716(9.708)	4.617(6.932)
M2	0.05	2.684(4.288)	4.465(**1.684**)	7.733(6.258)	**1.014**(3.343)
	0.25	**3.609(3.985)**	6.929(5.917)	11.670(11.963)	8.640(24.409)
	0.50	**9.072**(11.561)	18.820(13.056)	25.303(13.502)	36.565(**6.240**)
	0.75	**31.397(58.095)**	51.404(64.778)	63.834(152.047)	68.384(106.082)
	0.95	132.283(**166.269**)	**127.967**(250.006)	214.595(559.565)	282.834(646.127)
M3	0.05	**0.012(0.018)**	0.190(0.105)	0.207(0.116)	0.051(0.075)
	0.25	**0.011(0.018)**	0.048(0.111)	0.045(0.092)	0.057(0.079)
	0.50	**0.012(0.018)**	0.042(0.088)	0.040(0.105)	0.044(0.072)
	0.75	**0.014(0.019)**	0.084(0.130)	0.094(0.141)	0.052(0.071)
	0.95	**0.017(0.023)**	0.263(0.234)	0.296(0.197)	0.083(0.126)

Table 1: Mean squared errors (MSEs) and standard deviations (SDs) of the quantiles of samples generated by our conditional diffusion models, WGANs, Sinkhorn GANs, and k-NN. The smallest MSEs and SDs for each quantile are highlighted in bold.

which requires samples from both $p_{X,Y,Z}(x,y,z)$ and $p_{X,Z}(x,z)p_{Y|Z}(y|z)$. Following the approach of Li et al. (2024), we use the 1-NN sampling algorithm to estimate the conditional distribution of $Y|Z$. For further details, see Algorithm 4 in Supplementary Materials.

Finally, we formalize the classifier-based CMI estimator; see Algorithm 5 in Supplementary Materials. Specifically, consider a data set V consisting of $2n$ i.i.d. samples $\{W_i := (X_i, Y_i, Z_i)\}_{i=1}^{2n}$ with $(X_i, Y_i, Z_i) \sim p_{X,Y,Z}(x,y,z)$. We divide V into two parts, V_1 and V_2, each containing n samples. Using Algorithm 4, we generate a new data set V' with n samples from V_1 and V_2. Assigning labels $l = 1$ to all samples in V_2 (positive samples drawn from $p_{X,Y,Z}(x,y,z)$) and $l = 0$ to all samples in V' (negative samples drawn from $p_{X,Z}(x,z)p_{Y|Z}(y|z)$). In this supervised classification task, we train a binary classifier using advanced models like XGBoost (Sen et al. 2017; Chen and Guestrin 2016) or deep neural networks (Goodfellow, Bengio, and Courville 2016). The classifier outputs the predicted probability $\alpha_m = P(l = 1|W_m)$ for a given sample W_m, which leads to an estimator of the likelihood ratio on W_m given by $\widehat{L}(W_m) = \alpha_m/(1-\alpha_m)$. Based on Equations (4) and (6), we can derive an estimator for $I(X;Y|Z)$,

$$\begin{aligned}\widehat{I}(X;Y|Z) &:= \widehat{d}_{\text{KL}}(p_{X,Y,Z}(x,y,z)||p_{X,Z}(x,z)p_{Y|Z}(y|z)) \\ &= d^{-1}\sum_{i=1}^{d}\log\widehat{L}(W_i^f) - \log\{d^{-1}\sum_{j=1}^{d}\widehat{L}(W_j^g)\},\end{aligned} \tag{7}$$

where $d = \lfloor n/3 \rfloor$ with $\lfloor t \rfloor$ being the largest integer not greater than t, W_i^f is a sample in V_f^{test}, and W_j^g is a sample in V_g^{test}, with V_f^{test} and V_g^{test} defined in Algorithm 5.

Computation of the p-Value

We calculate the p-value based on the score-based conditional diffusion models to make informed decisions regarding the hypothesis testing. Specifically, by Algorithm 2, we independently draw pseudo samples $\widehat{X}_i^{(b)} \sim \widehat{p}(\cdot|Z_i)$ for each i across $b = 1, \ldots, B$, where $\widehat{p}(\cdot|Z_i)$ is the conditional density of $\widehat{X}$ given Z_i and B is the number of repetitions. Conditional on $\boldsymbol{Z}$, all $\widehat{\boldsymbol{X}}^{(b)} := (\widehat{X}_1^{(b)}, \ldots, \widehat{X}_n^{(b)})^T$ are independent of $\boldsymbol{Y}$ and also $\boldsymbol{X}$. We denote the CMI estimator of $I(X;Y|Z)$ based on $(\widehat{\boldsymbol{X}}^{(b)}, \boldsymbol{Y}, \boldsymbol{Z})$ as $\widehat{\text{CMI}}^{(b)}$ and denote the estimator based on $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ as $\widehat{\text{CMI}}$. According to Theorem 1 in Mukherjee, Asnani, and Kannan (2020), $\widehat{I}(X;Y|Z)$ is a consistent estimator of $I(X;Y|Z)$. We calculate the p-value using Equation (1) by substituting $\boldsymbol{T}(\boldsymbol{X}^{(b)}, \boldsymbol{Y}, \boldsymbol{Z})$ and $\boldsymbol{T}(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ with $\widehat{\text{CMI}}^{(b)}$ and $\widehat{\text{CMI}}$, respectively. The pseudo code is summarized in Algorithm 3. In the next section, we will prove that our test asymptotically achieves a valid control of type I error.

Theoretical Results

In this section, we present our main theoretical results, with all proofs deferred to Supplementary Materials. Denote $p^{(n)}(\cdot|\boldsymbol{Z}) := \prod_{i=1}^n p(\cdot|Z_i)$ and $\widehat{p}^{(n)}(\cdot|\boldsymbol{Z}) := \prod_{i=1}^n \widehat{p}(\cdot|Z_i)$. In Theorem 2, we bound the excess type I error conditionally on $\boldsymbol{Y}$ and $\boldsymbol{Z}$ by the total variation distance between $\widehat{p}^{(n)}(\cdot|\boldsymbol{Z})$ and $p^{(n)}(\cdot|\boldsymbol{Z})$.

Theorem 2. *Assume $H_0 : X \perp\!\!\!\perp Y|Z$ is true. For any significance level $\alpha \in (0,1)$, the p-value obtained from Algorithm 3 satisfies*

$$P(p \le \alpha|\boldsymbol{Y}, \boldsymbol{Z}) \le \alpha + d_{TV}(p^{(n)}(\cdot|\boldsymbol{Z}), \widehat{p}^{(n)}(\cdot|\boldsymbol{Z})).$$

An immediate implication of Theorem 2 is that the type I error rate can be unconditionally controlled as

$$P(p \le \alpha|H_0) \le \alpha + \mathbb{E}[d_{\text{TV}}(p^{(n)}(\cdot|\boldsymbol{Z}), \widehat{p}^{(n)}(\cdot|\boldsymbol{Z}))].$$

Then applying Theorem 1 to this inequality yields the following Corollary 1, which shows that our CI testing procedure can asymptotically control the type I error at level α.

Algorithm 3: Conditional diffusion models based conditional independence testing (CDCIT)

Input: Dataset $\mathcal{D}_{\mathrm{T}} = (\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ consisting of n i.i.d. samples from $p_{X,Y,Z}(x, y, z)$ and unlabelled dataset $\mathcal{D}_{\mathrm{U}} = (\boldsymbol{X}^{\mathrm{U}}, \boldsymbol{Z}^{\mathrm{U}})$ consisting of N i.i.d. samples from $p_{X,Z}(x, z)$.
Parameter: The number of repetitions B; the significance level α.
Output: Accept $H_0 : X \perp\!\!\!\perp Y|Z$ or $H_1 : X \not\perp\!\!\!\perp Y|Z$.

1: Use Algorithm 5 to obtain $\widehat{\mathrm{CMI}}$ based on $\mathcal{D}_{\mathrm{T}}$.
2: Use Algorithm 1 to obtain the score network $\widehat{s}$ based on $\mathcal{D}_{\mathrm{U}}$.
3: $b = 1$.
4: **while** $b \leq B$ **do**
5: For each $i \in \{1, \ldots, n\}$, given Z_i, produce $\widehat{X}_i^{(b)}$ using Algorithm 2 with Z_i and $\widehat{s}$ as input. Let $\widehat{\boldsymbol{X}}^{(b)} := (\widehat{X}_1^{(b)}, \ldots, \widehat{X}_n^{(b)})^T$.
6: Use Algorithm 5 to obtain $\widehat{\mathrm{CMI}}^{(b)}$ based on $(\widehat{\boldsymbol{X}}^{(b)}, \boldsymbol{Y}, \boldsymbol{Z})$.
7: $b = b + 1$.
8: **end while**
9: Compute p-value: $p := \big[1 + \sum_{b=1}^{B} \mathbf{I}\big\{\widehat{\mathrm{CMI}}^{(b)} \geq \widehat{\mathrm{CMI}}\big\}\big]/(1 + B)$.
10: **if** $p \geq \alpha$ **then**
11: Accept $H_0 : X \perp\!\!\!\perp Y|Z$.
12: **else**
13: Accept $H_1 : X \not\perp\!\!\!\perp Y|Z$.
14: **end if**

Corollary 1. *Assume* $n \cdot N^{-\Gamma_1(k,\alpha)} \cdot (\log N)^{\Gamma_2(k,\alpha)} = o(1)$. *Under assumptions in Theorems 1 and 2, the p-value obtained from Algorithm 3 satisfies*

$$P(p \leq \alpha | H_0) \leq \alpha + o(1).$$

Remark 1. *Since* $\log N = O(N^{\delta})$ *for any* $\delta > 0$, *the sample size assumption in Corollary 1 implies that the sample size* N *in* $\mathcal{D}_U$ *needs to satisfy* $N \gg n^{1/(\Gamma_1(k,\alpha)-\delta)}$ *for any sufficiently small* $\delta > 0$ *in order to asymptotically control the type I error.*

Synthetic Data Analysis

We evaluate our method, CDCIT, on synthetic datasets, and compare it with the seven state-of-the-art (SOTA) methods: GCIT (Bellot and van der Schaar 2019), NNSCIT (Li et al. 2023), the classifier-based CI test (CCIT) (Sen et al. 2017), the kernel-based CI test (KCIT) (Zhang et al. 2011), LPCIT (Scetbon, Meunier, and Romano 2022), DGCIT (Shi et al. 2021), and NNLSCIT (Li et al. 2024). We set the number of repetitions B to 100 and the significance level α to 0.05. We report the type I error rate and the testing power under H_1 for all methods in each experiment. All the results are presented as an average over 100 independent trials. We present additional simulation studies, the detailed training parameter settings for our CDCIT and the real data analysis in Supplementary Materials.

Scenario I: the post-nonlinear model. The first synthetic dataset is generated using the post-nonlinear model similar to those in Zhang et al. (2011), Bellot and van der Schaar (2019), Scetbon, Meunier, and Romano (2022), and Li et al. (2024). Specifically, the triples (X, Y, Z) under H_0 and H_1 are generated using the following models:

$$\begin{aligned} H_0 : X &= f_1(\overline{Z} + 0.25 \cdot \epsilon_x), \\ Y &= f_2(\overline{Z} + 0.25 \cdot \epsilon_y), \\ H_1 : X &= f_1(\overline{Z} + 0.25 \cdot \epsilon_x) + 0.5 \cdot \epsilon_b, \\ Y &= f_2(\overline{Z} + 0.25 \cdot \epsilon_y) + 0.5 \cdot \epsilon_b, \end{aligned} \tag{8}$$

where $\overline{Z}$ is the sample mean of $Z = (z_1, \ldots, z_{d_z})$, all z_l in Z, ϵ_x, ϵ_y and ϵ_b are i.i.d. samples generated from the standard Gaussian distribution, and functions f_1 and f_2 are randomly sampled from the set $\{x, x^2, x^3, \tanh(x), \cos(x)\}$, and d_z represents the dimension of Z.

Scenario II: the mixed continuous and discrete conditioning-set model. The conditioning variable set $Z = (z_1, \ldots, z_{d_z})$ is mixed-type, consisting of $\lfloor d_z/2 \rfloor$ continuous variables $(z_1, \ldots, z_{\lfloor d_z/2 \rfloor})$ and $d_z - \lfloor d_z/2 \rfloor$ discrete variables $(z_{\lfloor d_z/2 \rfloor + 1}, \ldots, z_{d_z})$. We use only $(z_1, z_2, \ldots, \lfloor 2 \cdot d_z/3 \rfloor)$ to generate X and Y under both H_0 and H_1 in the true model. Specifically,

$$\begin{aligned} H_0 : X &= \frac{1}{\lfloor 2 \cdot d_z/3 \rfloor} \sum_{i=1}^{\lfloor 2 \cdot d_z/3 \rfloor} z_i + 0.33 \cdot \epsilon_x, \\ Y &= \frac{1}{\lfloor 2 \cdot d_z/3 \rfloor} \sum_{i=1}^{\lfloor 2 \cdot d_z/3 \rfloor} z_i + 0.33 \cdot \epsilon_y, \\ H_1 : X &= \frac{1}{\lfloor 2 \cdot d_z/3 \rfloor} \sum_{i=1}^{\lfloor 2 \cdot d_z/3 \rfloor} z_i + 0.33 \cdot \epsilon_b, \\ Y &= \frac{1}{\lfloor 2 \cdot d_z/3 \rfloor} \sum_{i=1}^{\lfloor 2 \cdot d_z/3 \rfloor} z_i + 0.33 \cdot \epsilon_b, \end{aligned} \tag{9}$$

where $z_1, \ldots, z_{\lfloor d_z/2 \rfloor} \overset{i.i.d.}{\sim} N(0, 1)$, $z_{\lfloor d_z/2 \rfloor + 1}, \ldots, z_{d_z} \overset{i.i.d.}{\sim} 2 \cdot \text{Bernoulli}(1, 0.5) - 1$, and ϵ_x, ϵ_y and ϵ_b all follow the standard Gaussian distribution.

For each experiment, 1000 samples are generated. We use $N = 500$ to train the conditional sampler and $n = 500$ to compute the test statistic in our CDCIT. We vary d_z, the dimension of Z, from 10 to 100. The results are shown in Figure 2. More results regarding N and n are provided in Figures 4 and 5 in Supplementary Materials.

We have the following observations. First, in both post-nonlinear and mixed models, our test controls type I error very well and achieves high power under H_1 as d_z increases. Second, NNSCIT has satisfactory performance in controlling type I error, but it loses power under H_1, especially when d_z exceeds 40 in the mixed model. Third, although C-CIT, KCIT and GCIT have adequate power under H_1, they have inflated type I errors in almost all scenarios. Fourth, DGCIT and NNLSCIT sometimes fail to control the type I error well, especially when $d_z \leq 20$. Fifth, LPCIT shows weak performance on both type I error and testing power.

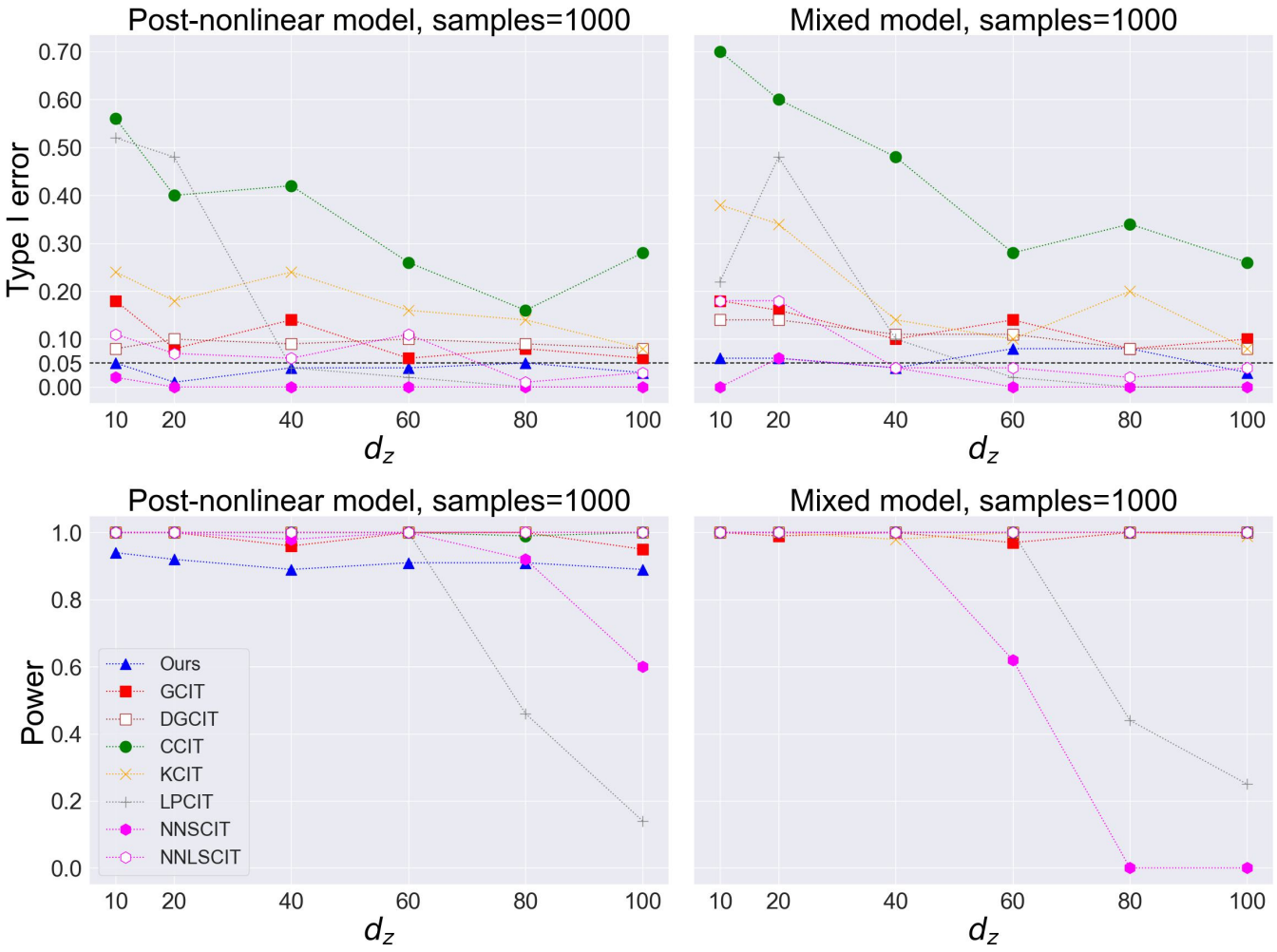

Figure 2: Comparison of the type I error (lower is better) and power (higher is better) of our method with seven SOTA methods on the post-nonlinear model (8) and mixed model (9) with varying dimension of Z. Under the mixed model, the power of our method, as well as those of DGCIT, CCIT, and NNLSCIT, stays consistently at 1 across different d_z.

Figure 7 in Supplementary Materials reports the timing performance of all considered methods for a single test. Our CDCIT is found to be highly computationally efficient even when dealing with large sample sizes and high-dimensional conditioning sets.

Conclusion

We introduce a novel CI testing procedure using the conditional diffusion models to approximate the distribution of $X|Z$. We have theoretically and empirically shown that the distribution of the generated samples is very close to the true conditional distribution. We use a computationally efficient classifier-based CMI estimator as the test statistic, which captures intricate dependence structures among variables. We demonstrate that our proposed test achieves valid control of the type I and type II errors. Furthermore, our test remains highly computationally efficient, even when dealing with high-dimensional conditioning sets. Our method has the potential to broaden the applicability of causal discovery in real-world scenarios, such as gene regulatory network, the identification of disease-associated genes, and intricate social networks, thereby aiding in the identification of relationships and patterns within complex systems.

Acknowledgments

Dr. Ziqi Chen's work was partially supported by National Natural Science Foundation of China (NSFC) (12271167 and 72331005) and Basic Research Project of Shanghai Science and Technology Commission (22JC1400800). We thank the anonymous reviewers for their helpful comments.

References

Arjovsky, M.; Chintala, S.; and Bottou, L. 2017. Wasserstein generative adversarial networks. In *International Conference on Machine Learning*, 214–223.

Belghazi, M. I.; Baratin, A.; Rajeshwar, S.; Ozair, S.; Bengio, Y.; Courville, A.; et al. 2018. Mutual information neural

estimation. In *International Conference on Machine Learning*, 531–540.

Bellot, A.; and van der Schaar, M. 2019. Conditional independence testing using generative adversarial networks. In *Advances in Neural Information Processing Systems*, volume 32.

Berrett, T. B.; Wang, Y.; Barber, R. F.; and Samworth, R. J. 2020. The conditional permutation test for independence while controlling for confounders. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(1): 175–197.

Candès, E.; Fan, Y.; Janson, L.; and Lv, J. 2018. Panning for gold: 'model-X' knockoffs for high dimensional controlled variable selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(3): 551–577.

Chen, T.; and Guestrin, C. 2016. Xgboost: A scalable tree boosting system. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

Cover, T. M.; and Thomas, J. A. 2012. *Elements of Information Theory*. John Wiley & Sons.

Dai, B.; Shen, X.; and Pan, W. 2022. Significance tests of feature relevance for a black-box learner. *IEEE Transactions on Neural Networks and Learning Systems*.

Dai, H.; Ng, I.; Luo, G.; Spirtes, P.; Stojanov, P.; and Zhang, K. 2024. Gene Regulatory Network Inference in the Presence of Dropouts: a Causal View. arXiv:2403.15500.

Dhariwal, P.; and Nichol, A. 2021. Diffusion models beat gans on image synthesis. In *Advances in Neural Information Processing Systems*, volume 34.

Doran, G.; Muandet, K.; Zhang, K.; and Schölkopf, B. 2014. A permutation-based kernel conditional independence test. In *Conference on Uncertainty in Artificial Intelligence*, 132–141.

Fu, H.; Yang, Z.; Wang, M.; and Chen, M. 2024. Unveil conditional diffusion models with classifier-free guidance: a sharp statistical theory. arXiv:2403.11968.

Fukumizu, K.; Gretton, A.; Sun, X.; and Schölkopf, B. 2007. Kernel measures of conditional dependence. In *Advances in Neural Information Processing Systems*, volume 20.

Genevay, A.; Peyré, G.; and Cuturi, M. 2018. Learning generative models with sinkhorn divergences. In *International Conference on Artificial Intelligence and Statistics*, 1608–1617.

Goodfellow, I.; Bengio, Y.; and Courville, A. 2016. *Deep learning*. MIT Press.

Hall, P.; Racine, J.; and Li, Q. 2004. Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association*, 99(468): 1015–1026.

Hall, P.; and Yao, Q. 2005. Approximating conditional distribution functions using dimension reduction. *The Annals of Statistics*, 33(3): 1404–1421.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33.

Ho, J.; Salimans, T.; Gritsenko, A.; et al. 2022. Video diffusion models. In *Advances in Neural Information Processing Systems*, volume 35.

Izbicki, R.; and Lee, A. B. 2017. Converting high-dimensional regression to high-dimensional conditional density estimation. *Electronic Journal of Statistics*, 11: 2800–2831.

Khera, A. V.; and Kathiresan, S. 2017. Genetics of coronary artery disease: discovery, biology and clinical translation. *Nature Reviews Genetics*, 18(6): 331–344.

Koller, D.; and Friedman, N. 2009. *Probabilistic graphical models: principles and techniques*. MIT Press.

Lauritzen, S. L. 1996. *Graphical models*, volume 17. Clarendon Press.

Li, C.; and Fan, X. 2020. On nonparametric conditional independence tests for continuous variables. *Wiley Interdisciplinary Reviews: Computational Statistics*, 12(3): e1489.

Li, S.; Chen, Z.; Zhu, H.; Wang, C.; and Wen, W. 2023. Nearest-neighbor sampling based conditional independence testing. In *AAAI Conference on Artificial Intelligence*, volume 37, 8631–8639.

Li, S.; Zhang, Y.; Zhu, H.; Wang, C.; Shu, H.; Chen, Z.; et al. 2024. K-nearest-neighbor local sampling based conditional independence testing. In *Advances in Neural Information Processing Systems*, volume 36.

Liu, M.; Katsevich, E.; Janson, L.; and Ramdas, A. 2022. Fast and powerful conditional randomization testing via distillation. *Biometrika*, 109(2): 277–293.

Mesner, O. C.; and Shalizi, C. R. 2020. Conditional mutual information estimation for mixed, discrete and continuous data. *IEEE Transactions on Information Theory*, 67(1): 464–484.

Mukherjee, S.; Asnani, H.; and Kannan, S. 2020. CCMI: Classifier based conditional mutual information estimation. In *Conference on Uncertainty in Artificial Intelligence*, 1083–1093.

Pearl, J. 1988. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan kaufmann.

Runge, J. 2018. Conditional independence testing based on a nearest-neighbor estimator of conditional mutual information. In *International Conference on Artificial Intelligence and Statistics*, 938–947.

Scetbon, M.; Meunier, L.; and Romano, Y. 2022. An asymptotic test for conditional independence using analytic kernel embeddings. In *International Conference on Machine Learning*, 19328–19346.

Sen, R.; Suresh, A. T.; Shanmugam, K.; Dimakis, A. G.; and Shakkottai, S. 2017. Model-powered conditional independence test. In *Advances in Neural Information Processing Systems*, volume 30.

Shi, C.; Xu, T.; Bergsma, W.; and Li, L. 2021. Double generative adversarial networks for conditional independence testing. *Journal of Machine Learning Research*, 22(285): 1–32.

Song, J.; Meng, C.; and Ermon, S. 2021. Denoising diffusion implicit models. In *International Conference on Learning Representations*.

Song, Y.; Sohl-Dickstein, J.; Kingma, D. P.; Kumar, A.; Ermon, S.; and Poole, B. 2021. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*.

Spirtes, P.; Glymour, C. N.; and Scheines, R. 2000. *Causation, prediction, and search*. MIT Press.

Su, L.; and White, H. 2008. A nonparametric Hellinger metric test for conditional independence. *Econometric Theory*, 24(4): 829–864.

Su, L.; and White, H. 2014. Testing conditional independence via empirical likelihood. *Journal of Econometrics*, 182(1): 27–44.

Wang, X.; Pan, W.; Hu, W.; Tian, Y.; and Zhang, H. 2015. Conditional distance correlation. *Journal of the American Statistical Association*, 110(512): 1726–1734.

Weglarczyk, S. 2018. Kernel density estimation and its application. *ITM Web of Conferences*, 23: 00037.

Yang, L.; Zhang, Z.; Song, Y.; et al. 2023. Diffusion models: A comprehensive survey of methods and applications. *ACM Computing Surveys*, 56(4): 1–39.

Zan, L.; Meynaoui, A.; Assaad, C. K.; et al. 2022. A conditional mutual information estimator for mixed data and an associated conditional independence test. *Entropy*, 24(9): 1234.

Zhang, K.; Peters, J.; Janzing, D.; and Schölkopf, B. 2011. Kernel-based conditional independence test and application in causal discovery. In *Conference on Uncertainty in Artificial Intelligence*, 804–813.

Zhu, Z.; Zheng, Z.; Zhang, F.; et al. 2018. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nature Communications*, 9(1): 1–12.