arXiv:2304.04183v1 [cs.LG] 9 Apr 2023

基于最近邻采样的条件独立性检验

Shuai Li[1], Ziqi Chen[1*], Hongtu Zhu[2],Christina DanWang[3],WangWen[4][1]华东师范大学统计学院，教育部统计与数据科学重点实验室，中国上海[2] 北卡罗来纳大学教堂山分校生物统计学、统计学、计算机科学与遗传学系，美国教堂山 [3] 纽约大学上海商学院，中国上海[4] 中南大学数学与统计学院，中国长沙
shuaili.cq@foxmail.com, zqchen@fem.ecnu.edu.cn, htzhu@email.unc.edu, christina.wang@nyu.edu, wangwen.a@foxmail.com

摘要

近期提出了条件随机化检验（CRT），用于检验给定随机变量 Z时，两个随机变量 X 和 Y是否条件独立。CRT 假设在原假设下，给定 Z 时 X 的条件分布已知，并将其与原始数据观测样本的分布进行比较。本文旨在通过最近邻采样提出一种新颖的 CRT 替代方法，无需假设给定 Z时 X 的分布具有确切形式。具体而言，我们利用计算效率较高的 1-最近邻来近似编码原假设的条件分布。随后，我们从理论上证明，生成样本的分布与真实条件分布之间的全变差距离很小。此外，我们采用基于分类器的条件互信息估计量作为检验统计量。作为一种经验性的基础信息论量，该检验统计量能够很好地捕捉条件依赖特征。我们表明，所提出的检验在控制第一类和第二类错误方面表现良好，同时计算速度非常快。最后，我们通过合成数据和真实数据分析展示了所提出检验的有效性。

引言

条件独立性检验（CIT）在统计学和机器学习中有着广泛的应用，包括因果推断（Spirtes 等人 2000；Pearl 2009；Cai、Li 和 Zhang 2022年）和图模型（Lauritzen 1996；Koller 和 Friedman 2009）这两个广为人知的例子。本文旨在开发一种灵活且快速的条件独立性检验方法。具体而言，我们考虑两个一元连续随机变量 X 和 Y，以及一组随机变量 $Z \in R^{d_Z}$，其维度 d_Z 可能趋于无穷大，且联合密度函数为 $p_{X,Y,Z}(x,y,z)$。基于n 个独立同分布（i.i.d）的副本$\{(X_i, Y_i, Z_i): i = 1, \ldots, n\}$，它们来自 (X, Y, Z)，我们关注检验两个随机变量 X 和 Y 在给定 Z的条件下是否条件独立；即，

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

其中，$\perp\!\!\!\perp$ 表示独立性。 Z 的高维性使条件独立性检验颇具挑战（Bellot 和 van der Schaar 2019；Shi 等人 2021）。我们提出的方法可以直接扩展到 X 和 Y为多元变量的情形。

近年来，人们提出了许多检验条件独立性的方法。例如，Candes 等人（2018）、Zhang 等人（2011）、Zhang 等人（2017）、Bellot 和 van der Schaar（2019）、Strobl、Zhang 和 Visweswaran（2019）、Berrett 等人（2020）、Shah 和 Peters（2020）、Shi 等人（2021）以及 Zhang 等人（2022）的方法。其中，Candes 等人（2018）提出的条件随机化检验（CRT）是最重要的方法之一，但 CRT 假设真实的条件分布 $p_{X|Z}$ 已知。给定 $\{Z_1, \ldots, Z_n\}$，可以独立抽取$X_i^{(m)} \sim p_{X|Z=Z_i}$，对 $m = 1$中的每个 i，从而使所有 $\boldsymbol{X}^{(m)} := (X_1^{(m)}, \ldots, X_n^{(m)})$ 与$\boldsymbol{X} := (X_1, \ldots, X_n)$ 以及 $\boldsymbol{Y} := (Y_1, \ldots, Y_n)$ 相互独立，其中 M是重复次数。因此，在原假设H_0：$X \perp\!\!\!\perp Y|Z$ 下，对于所有m，都有 $(X^{(m)}, Y, Z) \overset{d}{=} (X, Y, Z)$，其中 $\overset{d}{=}$ 表示分布相等。$(X^{(m)}, Y, Z)$ 与 (X, Y, Z) 之间的较大差异可视为反对 H_0的有力证据。从统计学角度，可以考虑检验统计量 $T(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$，并近似其 p值，方法是

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

其中 $I(\cdot)$ 是示性函数。在 H_0 下，p-值有效，并且对于任意 $\alpha \in (0, 1)$，$P(p \leq \alpha | H_0) \leq \alpha$ 成立。

由于 $p_{X|Z}$ 在实践中鲜为人知，人们基于对 $p_{X|Z}$ 的不同近似方法开发了多种方法。例如，Bellot 和 van der Schaar（2019）利用 Wasserstein 生成对抗网络（WGAN，Arjovsky、Chintala 和 Bottou，2017）近似 $p_{X|Z}$，开发了生成式条件独立性检验（GCIT）。令 $\widehat{p}_{X|Z}$ 为基于 WGAN 的 $p_{X|Z}$ 估计量。理论上，正如 Bellot 和 van der Schaar（2019）所示，他们的 GCIT 检验超过目标水平 α 的第一类错误，其上界为 $E\{d_{TV}(p_{\boldsymbol{X}|\boldsymbol{Z}}, \widehat{p}_{\boldsymbol{X}|\boldsymbol{Z}})\}$，其中 d_{TV} 表示全变差距离。然而，图 1 显示，$\widehat{p}_{X|Z}$ 对 $p_{X|Z}$ 的近似效果很差，

*通讯作者：Ziqi Chen。版权所有 © 2023年，人工智能促进协会 (www.aaai.org)。保留所有权利。

在两个相对简单的模拟设置中。因此，正如合成数据分析所示，GCIT 检验的第一类错误率偏高。最近，Shi 等人（2021）提出使用 Sinkhorn GAN（Genevay、Peyr´e 和 Cuturi，2018）近似 $p_{X|Z}$。如图 1 所示，我们发现，Sinkhorn GAN 在这两个相对简单的模拟设置中的表现也很差。

在 CRT 中，检验统计量的选择对于获得足够的统计功效以及控制第一类错误至关重要，但这一问题尚未得到仔细研究。例如，Bellot 和 van der Schaar（2019）提出考虑多种检验统计量，包括最大均值差异（MMD）、Pearson 相关系数（PCC）、距离相关（DC）和 Kolmogorov-Smirnov 距离（KS），但对于如何在不同情境下恰当地选择检验统计量，人们知之甚少。此外，仅衡量 X 与 Y 之间依赖关系的检验统计量，在 H_1 下可能会导致第一类错误率升高和／或功效不足。我们考虑两种情境：(i) 简单马尔可夫链 $X \to Z \to Y$；(ii) $X \to Z \leftarrow Y$，其中连接两个随机变量的直接箭头表示直接因果关系。在这两种情境下，仅衡量 X 与 Y 之间依赖关系的检验统计量，在检验 H_0 对 H_1 时都会增加第一类错误率和／或降低统计功效。因此，需要使用能够捕捉条件依赖关系的检验统计量。本文考虑 (X, Y, Z) 的条件互信息（CMI），记为 $I(X;Y|Z)$，以及其经验版本（Mukherjee、Asnani 和 Kannan 2020），因为它为条件依赖关系提供了强有力的理论保证，即 $I(X;Y|Z) = 0 \iff X \perp\!\!\!\perp Y|Z$（Cover 和 Thomas 2012）。CMI 已广泛应用于因果学习（Hlinka 等，2013）、图模型（Liang 和 Wang，2008）以及特征选择（Fleuret，2004）。然而，经验 CMI 的计算可能十分困难，尤其是在处理高维 Z 时。

本文提出了一种基于 1-近邻采样策略（NNSCIT）的新型 CIT 方法，用于模拟来自与真实密度 $p_{X|Z}$ 近似相近的分布的样本。Fix 和 Hodges（1951）最早提出的最近邻采样，已广泛应用于密度估计、分类和回归问题（Silverman 2018；Cover 和 Hart 1967；Devroye 等，1994）。最近，Sen 等人（2017）使用最近邻自助法从 $X \perp\!\!\!\perp Y|Z$ 下的联合分布 (X, Y, Z) 中生成样本。与 GAN 相比，1 近邻（ 1-NN ）不仅计算效率高，而且在近似质量方面也表现出优势。

我们的四项主要贡献如下。首先，我们提出使用1-近邻方法，从 X 给定 Z 的近似条件分布中生成样本。与 WGAN（Bellot 和 van der Schaar2019）和Sinkhorn GAN（Shi 等人 2021）相比，1-近邻在计算上高效得多。从理论上讲，我们证明，就全变差距离而言，由 1-近邻生成的样本的分布与真实条件分布非常接近。其次，我们将 $I(X;Y|Z)$ 作为检验统计量，并使用近期基于分类器的方法（Mukherjee、Asnani 和 Kannan 2020）。第三，对于由 1-NN 生成的伪样本 $\widetilde{X}^{(m)}(m = 1, \ldots\ldots, M)$，我们提出可用 $I(\widetilde{X}^{(m)};Y)$ 替代 $I(\widetilde{X}^{(m)};Y|Z)$，以加快计算，因为 $I(\widetilde{X}^{(m)};Y|Z)$ 的估计计算量很大，尤其是在 Z 的维数较高时。第四，我们提出的检验不仅在渐近意义下能有效控制第一类错误，而且在数值研究中优于所有具有竞争力的检验。

算法 1：1-近邻采样算法 (1-NN(V_1,V_2,n))输入：数据集 V_1 和 V_2，二者的样本量为 n，且$V = V_1 \cup V_2$ 由来自p_X、z的 $2n$ 独立同分布样本组成。~输出：对于V_2中的每个 Z-坐标，根据 $X|Z$ 生成 X 。1：令 $U_0 = \emptyset$。2：对于 V_2 中的(X, Z)执行~ ~ ~3：转到 V_1，以找到样本(X, Z)，使得 Z 是 Z 关于 l_2 范数的1-最近邻。~4：$U_0 = U_0 \cup \{X\}$。5：结束循环 6：返回 U_0

1-近邻采样算法

在本节中，我们介绍1近邻采样算法，以及其理论和实证结果；这些结果表明，所生成样本的分布与真实条件分布非常接近。

从 $p_{X|Z}$（$x|z$） 进行 1近邻采样

我们有两个数据集 V_1 和 V_2，二者的样本量均为 n，其中 $V = V_1 \cup V_2$ 由来自分布$p_X(z,x)$的 $2n$ 独立同分布样本组成。给定 V_2中的所有 Z 坐标，算法1给出了生成数据集 U_0 的过程，该数据集由 n 个样本组成，模拟从$p_{X|Z}(x|z)$生成的样本。具体而言，对于 V_2 中的每个 Z 坐标，我们根据 l_2 范数下的 ~Z 坐标，在 V_1 中搜索最近邻 (X, Z)，然后将 X 添加到 U_0 中。当 V 是一个包含来自分布$p_{X,Y,Z}(x, y, z)$的样本的集合时，算法 1 在忽略 Y 坐标的情况下仍然适用。

理论结果

对于 Z 中的一个坐标，我们证明，在全变差距离意义下，算法 1 中生成的 ~ 的分布非常接近真实的条件分布。在介绍我们的理论结果之前，我们首先介绍 Cover 和Hart（1967）的引理 1，该引理指出，当训练样本量 n 趋于无穷大时， Z 的最近邻几乎必然收敛于 Z 。

引理 1. 设 Z *and* $Z_1, Z_2, \ldots, Z_n$ 是服从 $p(z)$ 的独立同分布随机变量。设 Z'_n 是集合 $\{Z_1, Z_2, \ldots, Z_n\}$ 中距离 Z 最近的邻居。那么，随着 n 趋于无穷，Z'_n 几乎必然收敛到 Z 。

接下来，我们介绍若干标准正则性条件，这些条件已由 Gao、Oh 和 Viswanath（2016）、Gao、Oh 和 Viswanath（2017）以及 Sen 等人（2017）提出。为简洁起见，可以省略下标。例如，可以用 $p(x|z)$ 代替 $p_{X|Z}(x|z)$。

关于 $p(x|z)$的平滑性假设： 假设 $p(x|z)$ 满足平滑性条件，这可视为对 x 关于 z 的费舍尔信息矩阵最大特征值有界这一条件的推广。

假设 1. 对于所有的 $z \in R^{d_z}$ 以及所有满足 $\|a-z\|_2 \leq \epsilon_1$ 的 a，都有 $0 \leq \lambda_{max}(I_a(z)) \leq \beta$，其中 $\beta > 0$，$\|\cdot\|_2$ 为 l_2 范数，广义曲率矩阵 $I_a(z) = (I_a(z)_{ij})$ 定义为

$$I_a(z)_{ij} = E\left(-\frac{\partial^2 \log p(x|\widetilde{z})}{\partial \widetilde{z}_i \partial \widetilde{z}_j}|_{\widetilde{z}=a} \middle| Z = z\right)$$
$$= \left(\frac{\partial^2}{\partial \widetilde{z}_i \partial \widetilde{z}_j} \int \log \frac{p(x|z)}{p(x|\widetilde{z})} p(x|z) dx\right)\Bigg|_{\widetilde{z}=a}.$$

关于 $p(z)$：

假设2。 概率密度函数 $p(z)$ 二阶连续可微，且 $p.d.f.$ $p(z)$关于 z 的 Hessian 矩阵 $H_p(z)$几乎处处满足$\|H_p(z)\|_2 \leq c_{d_z}$，其中 c_{d_z} 仅依赖于 d_z。

~给定 Z，令 X 表示由 1近邻 产生的样本，使得 ~ 中的 $X = X'_n$ 是样本 (X'_n, Z'_n) 在 V_1 中的 X 坐标，其中 Z'_n 是 Z 的最近邻。毫无疑问，$X \sim p(x|Z'_n)$。令 $\widehat{p}(x|Z) := p(x|Z'_n)$。对于定义在同一概率空间上的任意两个分布 P_1 和 P_2，P_1 与 P_2 之间的全变差距离定义为 $d_{TV}(P_1, P_2) = \sup_{A \subset \Omega} |P_1(A) - P_2(A)|$，其中上确界取遍样本空间 Ω 的所有可测子集。下面给出定理，其证明留在补充材料中。

定理 2。 在假设 *1* 和 *2*下，有$d_{TV}(p(x|Z)\ p(x|Z'_n)) = d_{TV}(p(x|Z)\ \widehat{p}(x|Z)) = o_p(1)$，当样本量 n 在 V_1 中趋于无穷时。

经验拟合优度

在本小节中，我们考察由 1近邻生成的样本的经验拟合优度表现。我们考虑以下两种情形。

情形 1。 $X \sim$ 均匀分布$[0,1]$ 和 Z 被假设为相互独立，其中 Z 服从均值向量为 $(0.7, 0.7, \ldots, 0.7)$、协方差矩阵为单位矩阵的 50 维多元高斯分布。$X|Z$ 的真实条件分布与 X 的相同。

场景 2。 设 $X = A_f^T Z + \epsilon$，其中 A_f 的各个条目从 $[0,1]$ 中随机均匀采样，然后归一化为单位 l_1 范数，并且 Z 从均值向量为 (0。7，0。7，.. .，0。7)、协方差矩阵为单位矩阵的 50 维多元高斯分布中生成。噪声变量 ϵ’ 独立地从均值为零、方差为 0.49 的正态分布中采样。

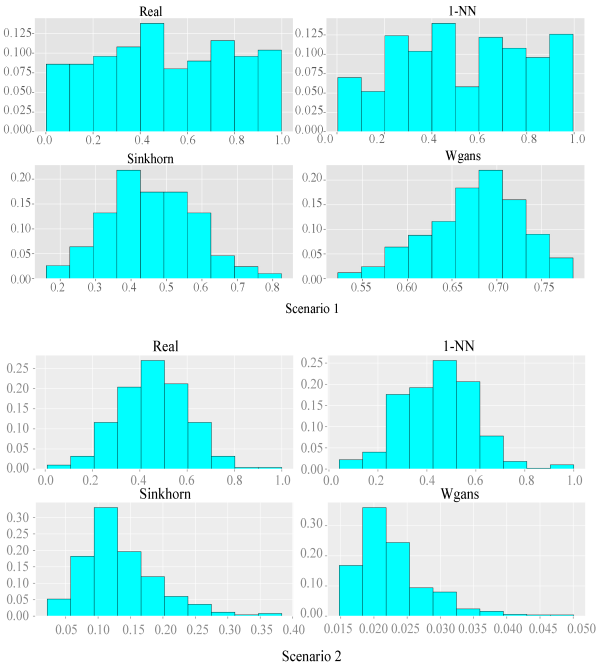


图 1：条件直方图。

对于两个模型中的每一个，我们生成 $n = 1000$ 个样本。随机选取 500 个样本作为训练数据集 V_1，其余样本作为测试数据集 V_2。对于 1近邻，我们通过 1近邻(V_1, V_2, 500)。生成 500 个伪样本。给定 V_2 中的每个 Z 坐标，我们还分别使用 WGAN 和 Sinkhorn GAN 生成伪样本 $\sim X$。图 1 展示了场景 1 和场景 2 中生成样本以及真实样本的条件直方图，所有样本均归一化到范围 $[0,1]$。可以观察到，1近邻能够较好地拟合条件密度，而 WGAN 和 Sinkhorn GAN 的表现较差。具体而言，WGAN 往往偏向于 0 或 1，而 Sinkhorn GAN 无法捕捉真实条件分布的特征。

基于最近邻采样的 CIT

本节中，我们介绍基于最近邻采样的 CIT（NNSCIT），并在算法 2 中给出计算 NNSCIT 及其 p 值的伪代码。此外，我们从理论上证明，所提出的检验能够有效控制第一类错误率。

所提出的条件独立性检验方法

我们的CIT 检验基于对 CMI$I(X;\ Y|Z) = I(X;\ Y, Z) - I(X;\ Z)$ 的近似，其中 $I(X;\ Y, Z)$ 和 $I(X;\ Z)$分别是 $(X; Y, Z)$ 的互信息和$(X; Z)$ 的互信息。我们将 CIT 统计- 量构造为 $I(X; Y|Z)$ 的基于分类器的 CMI 估计量（CCMI, Mukherjee、Asnani 和 Kannan，2020），其定义为

$$\widehat{I}(X; Y|Z) = \widehat{I}(X; Y, Z) - \widehat{I}(X; Z), \qquad (2)$$

$\widehat{\ }$ $\widehat{\ }$其中，$I(X;Y,Z)$（或 $I(X;Z)$）是 $I(X;Y,Z)$（或 $I(X;Z)$）的基于分类器的估计量。根据 Mukherjee、Asnani 和Kannan（2020）中的定理1，$I(X;Y|Z)$是 $I(X;Y|Z)$的一致估计量。此外，从在 Z条件下的 1近邻生成样本 $\widetilde{X}^{(m)}(m=1,\ldots,M)$，我们可以证明，对于所有$m$，$\widehat{I}(\widetilde{X}^{(m)};Y|Z)-I(\widetilde{X}^{(m)};Y|Z)$ 收敛于零。

在不失一般性的情况下，我们讨论如何近似 $I(X;Z)=D_{KL}(p_{X,Z}(x,z)||p_X(x)p_Z(z))$，其中$p_X$，$z(x,z)$是 (X,Z) 的联合密度，而 $p_X(x)$和$p_Z(z)$分别是 X 和 Z的边缘密度。此外，$D_{KL}(F||G)=\int f(x)\log(f(x)/g(x))dx$ 是两个分布函数 F 和 G之间的Kullback-Leibler（KL）散度，其密度函数分别由$f(x)$ 和 $g(x)$给出。$D_{KL}(F||G)$ 的 Donsker-Varadhan（DV）表示为

$$\sup_{s\in\mathcal{S}}\left[E_{x\sim f}s(x)-\log\{E_{x\sim g}\exp(s(x))\}\right], \tag{3}$$

其中，函数类 $\mathcal{S}$ 包含式（3）中所有期望有限的函数。式（3）中的最优函数为 $s^*(x)=\log(f(x)/g(x))$（Belghazi 等，2018），从而得出

$$D_{KL}(F||G)=E_{x\sim f}\log\left\{\frac{f(x)}{g(x)}\right\}-\log\left[E_{x\sim g}\left\{\frac{f(x)}{g(x)}\right\}\right]. \tag{4}$$

遵循 Mukherjee、Asnani 和 Kannan (2020)，我们使用分类器双样本原理（Lopez-Paz 和 Oquab 2017）来估计似然比 $L(x)=f(x)/g(x)$，具体如下。具体而言，我们考虑 n 独立同分布样本 $\{X_i^f\}_{i=1}^n$，其分布为 $X_i^f\sim f(x)$，以及 d 独立同分布样本 $\{X_j^g\}_{j=1}^d$，其分布为 $X_j^g\sim g(x)$。对于所有 X_i^f，我们将 $y_i^f=1$ 标记为一类；对于所有 X_j^g，我们将 $y_j^g=0$ 标记为另一类。在这项监督分类任务中，我们使用深度神经网络训练二元分类器。对于给定样本 X_l，分类器会输出预测概率 $\alpha_l=Pr(y=1|X_l)$，从而得到 $X_l\widehat{\ }$ 上似然比的估计量 $L(X_l)=\alpha_l/(1-\alpha_l)$。因此，由$\widehat{\ }$(4)可知，KL 散度的估计量 $D_{KL}(F||G)$ 为

$$n^{-1}\sum_{i=1}^n\log\widehat{L}(X_i^f)-\log\left\{d^{-1}\sum_{j=1}^d\widehat{L}(X_j^g)\right\}.$$

由于互信息是 KL 散度的一种特殊情况，我们得到 $I(X;Z)$ 的估计量 $I(X;Z)$，以及 $I(X;Y,Z)$ 的估计量。

遵循 CRT 的思路，我们的 CIT 方法的 p值可表示为

$$p=\frac{1+\sum_{m=1}^M I\left(\widehat{I}(\widetilde{X}^{(m)};Y|Z)\geq\widehat{I}(X;Y|Z)\right)}{1+M}. \tag{5}$$

在引理3 中，我们表明，基于(5) 的检验的超额第一类错误由 $p_{\boldsymbol{X}|\boldsymbol{Z}}(\cdot|\boldsymbol{Z})$ 与 $\widehat{p}_{\boldsymbol{X}|\boldsymbol{Z}}(\cdot|\boldsymbol{Z})$ 之间的全变差距离所界定。根据定理 2，我们进一步得到 $P(p\leq\alpha|H_0)\leq\alpha+o(1)$。因此，我们的条件独立性检验（CIT）方法的超额第一类错误保证趋于零作为 $n\to\infty$。应训练两个基于深度神经网络的二元分类器，以为每个$\sim m$获取 $\widehat{I}(\widetilde{X}^{(m)};Y|Z)$。结合 $I(X;Y|Z)$，应训练 $2(M+1)$ 个二元分类神经网络，以计算公式（5）中的 p-值。当 M 较大时，计算量极大且耗时，尤其是在 Z 的维数较高的情况下。

在 (5) 中，我们进一步建议使用根据以下理由通过 Mesner 和 Shalizi (2020) 的方法计算得到的 $\widehat{I}(\widetilde{X}^{(m)};Y)$，而非使用 $\widehat{I}(\widetilde{X}^{(m)};Y|Z)$。首先，与 $\widehat{I}(\widetilde{X}^{(m)};Y|Z)$ 相比，$\widehat{I}(\widetilde{X}^{(m)};Y)$ 的计算速度非常快。其次，$\widetilde{X}^{(m)}$ 是在给定 Z 的条件下通过 1-NN 生成的，因此我们有 $I(\widetilde{X}^{(m)};Y|Z)=0$；而 $\widetilde{X}^{(m)}$ 和 Y 可能通过 Z 共享信息，即 $I(\widetilde{X}^{(m)};Y)\geq I(\widetilde{X}^{(m)};Y|Z)=0$。根据 $\widehat{I}(\widetilde{X}^{(m)};Y|Z)$ 和 $\widehat{I}(\widetilde{X}^{(m)};Y)$ 的一致性，我们得出结论：将 $\widehat{I}(\widetilde{X}^{(m)};Y|Z)$ 替换为 $\widehat{I}(\widetilde{X}^{(m)};Y)$，可以更好地控制我们的 CIT 方法发生第一类错误的概率。因此，我们提出了 (5) 的一个简单替代形式，用于计算 p值，如下所示：

$$p=\frac{1+\sum_{m=1}^M I\left(\widehat{I}(\widetilde{X}^{(m)};Y)\geq\widehat{I}(X;Y|Z)\right)}{1+M}. \tag{6}$$

由于 $\widetilde{X}_i^{(m)}$ s 是通过 1近邻采样策略生成的，我们将该检验称为 NNSCIT。公式 (6) 为我们提出的条件独立性检验方法奠定了基础，其伪代码已总结于算法 2 中。

我们介绍如何获得 $\widehat{I}(\widetilde{X}^{(m)};Y)$。具体而言，$\sim$ $\widetilde{\ }$给定独立同分布样本 $\{(X_i,Y_i)\}_{i=1}^n$，其中 $(X_i,Y_i)\sim p_{\widetilde{X},Y}$。$\widetilde{\ }$令 $\rho_{k,i}/2$ 表示点 (X_i,Y_i) 到其第 k 个最近邻的 l_∞-距离。定义

$$n_{\widetilde{X},i}=|\{\widetilde{X}_j:|\widetilde{X}_i-\widetilde{X}_j|\leq\rho_{k,i}/2\,,j\neq i\}|,$$

其中 $|A|$ 是集合 A 中元素的数量。类似地，定义 $n_{Y,i}$。对于每个 i，我们定义

$$\delta_i=\psi(k)-\psi(n_{\widetilde{X},i})-\psi(n_{Y,i})+\psi(n),$$

其中 $\psi(x):=d\log\Gamma(x)/dx$ 是 digamma 函数。因此，我们有

$$\widehat{I}(\widetilde{X};Y)=\max\left\{\frac{1}{n}\sum_{i=1}^n\delta_i,0\right\}. \tag{7}$$

根据 Mesner 和 Shalizi$\widehat{\ }$ $\sim$ $\widetilde{\ }$(2020) 中的定理 3.1 和 3.2 可知，$I(X;Y)$ 是 $I(X;Y)$ 的一致估计量。

最后(6)我们讨论为何不能将$I(X;Y|Z)$替换为$\sim I(X;Y)\widehat{\ }$。我们可以考虑如下近似 p-值：

$$p=\frac{1+\sum_{m=1}^M I\left\{\widehat{I}(\widetilde{X}^{(m)};Y)\geq\widehat{I}(X;Y)\right\}}{1+M}, \tag{8}$$

从而得到另一个 CRT 检验。令 $\widehat{c}_\alpha$ 为分布 $\widehat{I}(\widetilde{X}^{(m)};Y)$ 的上 α分位数。给定显著性水平 α，式 (6) 和式 (8) 的拒绝域分别为

算法 2：基于最近邻采样的条件独立性检验（NNSCIT）输入：来自 p_X、$_Y$、$_Z$ 的 n 个独立同分布样本组成的数据集 U。参数：重复次数 M；互信息估计中的邻居阶数 k；显著性水平 α。输出：接受 H_0：$X \perp\!\!\!\perp Y|Z$，或 H_1：$X \not\perp\!\!\!\perp Y|Z$。1：将 U 随机划分为两个不相交的部分：$U_1 := \{X_{train}, Y_{train}, Z_{train}\}$，样本量为 $n - \lfloor n/3 \rfloor$；以及 $U_2 := \{X_{test}, Y_{test}, Z_{test}\}$，样本量为$\lfloor n/3 \rfloor$。2：$m = 1$。3：当 $m \le M$ 时，执行 4：从 U_1 中随机抽取 $\lfloor n/3 \rfloor$ 个样本，得到V_1。5：使用算法 1 中的 1近邻(V_1,U_2, $\lfloor n/3 \rfloor$) 生成 U_0^m：$= \{\widetilde{X}^{(m)}\}$。6：根据式 (7) 计算 $I^{(m)} := \widehat{I}(\{\widetilde{X}^{(m)}\}; \{Y_{test}\})$。7：$m = m + 1$。8：结束循环。9：根据式 (2) 计算 $I := I(\{X_{test}\}; \{Y_{test}\}|\{Z_{test}\})$。10：计算 p 值：$p := \frac{1+\sum_{m=1}^{M} I\{I^{(m)} \ge I\}}{1+M}$。11：如果 $p \ge \alpha$，则 12：接受 H_0：$X \perp\!\!\!\perp Y|Z$。13：否则 14：接受 H_1：$X \not\perp\!\!\!\perp Y|Z$。15：结束条件判断

$\widehat{}$ $\widehat{}$分别由 $\{I(X; Y|Z) > \hat{c}_\alpha\}$ 和 $\{I(X;Y) > \hat{c}_\alpha\}$给出。在 H_1下，$I(X;Y|Z)$ 应当偏离零。直观而言，$\widehat{}$拒绝域为 $\{I(X;Y|Z) > \hat{c}_\alpha\}$ 的检验比拒绝域为 $\{I(X;Y) > \hat{c}_\alpha\}$的检验更$\widehat{}$容易接受 H_1。例如，考虑 $X \to Z \leftarrow Y$。这一关系表明$X \not\perp\!\!\!\perp Y|Z$（$H_1$ 成立），但 X 和 Y 可能相互独立。$\widehat{}$因此，拒绝域 $\{I(X;Y|Z) > \hat{c}_\alpha\}$ 可能$\widehat{}$检测到 H_1，但 $\{I(X;Y) > \hat{c}_\alpha\}$ 可能无法做到。也就是说，在 H_1下，使用 (6) 的检验通常比使用 (8) 的检验更有力。再考虑另一种特殊情况：当$X \perp\!\!\!\perp Z$时，我们得到 $I(X;Y|Z) \ge I(X;Y)$。根据 $I(X;Y|Z)$ 和 $I(X;Y)$ 的一致性，用 $I(X;Y|Z)$ 替换$\widehat{}$(8) 中的 $I(X;Y)$，会提高在 H_1下的检验力。也就是说，在 H_1下，(6) 比 (8) 具有更高的检验力。对于 $Y \perp\!\!\!\perp Z$，我们也可以得出相同的结论。

理论结果

在本小节中，我们基于 (5) 和 (6) 给出 NNSCIT 的理论结果。我们引入以下记号。一般性不失，令 $U_2 := \{(X_1, Y_1, Z_1) \ldots, (X_{n_1}, Y_{n_1}, Z_{n_1})\}$ 在算法 2 中，且$n_1 = \lfloor n/3 \rfloor$，其中 $\lfloor x \rfloor$ 是小于或等于 x的最大整数。定义 $\boldsymbol{X} := (X_1, X_2, \ldots, X_{n_1})$、$\boldsymbol{Y} := (Y_1, Y_2, \ldots, Y_{n_1})$ 和 $\boldsymbol{Z} := (Z_1, Z_2, \ldots, Z_{n_1})$。记$\sim$$P(\cdot|\boldsymbol{Z}) := p(\cdot|Z_1) \times \ldots \times p(\cdot|Z_{n_1})$，以及 $P(\cdot|\boldsymbol{Z}) := \widehat{p}(\cdot|Z_1) \times \ldots \times \widehat{p}(\cdot|Z_{n_1})$。假设 $\widetilde{\boldsymbol{X}}^{(m)} := (\widetilde{X}_1^{(m)}, \ldots, \widetilde{X}_{n_1}^{(m)})\widehat{}$是根据 $P(\cdot|\boldsymbol{Z})$ 抽样得到的，其中 $m = 1, \ldots, M$。令 $T(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}) := \widehat{I}(X; Y|Z)$，且 $T(\widetilde{\boldsymbol{X}}^{(1)}, \boldsymbol{Y}, \boldsymbol{Z}) := \widehat{I}(\widetilde{X}^{(1)}; Y|Z)$ $\ldots, T(\widetilde{\boldsymbol{X}}^{(M)}, \boldsymbol{Y}, \boldsymbol{Z}) := \widehat{I}(\widetilde{X}^{(M)}; Y|Z)$。

$\widehat{}$令 $\boldsymbol{X}_F$ 为从 $P(\cdot|\boldsymbol{Z})$ 中额外独立抽取的一个副本，且与 $\boldsymbol{Y}$ 和 $\boldsymbol{X}$、$\widetilde{\boldsymbol{X}}^{(1)}$、.....、$\widetilde{\boldsymbol{X}}^{(M)}$相互独立。在 H_0：$X \perp\!\!\!\perp Y|Z$ 条件下，给定 $\boldsymbol{Y}$ 和 $\boldsymbol{Z}$，$\boldsymbol{X}$与 ($\widetilde{\boldsymbol{X}}^{(1)}$、.....、$\widetilde{\boldsymbol{X}}^{(M)}$) 相互独立，且 $\widetilde{\boldsymbol{X}}_F$ 与 ($\widetilde{\boldsymbol{X}}^{(1)}$、.....、$\widetilde{\boldsymbol{X}}^{(M)}$) 相互独立。因此，我们有

$$d_{TV}\{((\boldsymbol{X}, \widetilde{\boldsymbol{X}}^{(1)}, \ldots, \widetilde{\boldsymbol{X}}^{(M)})|\boldsymbol{Y}, \boldsymbol{Z}), ((\widetilde{\boldsymbol{X}}_F, \widetilde{\boldsymbol{X}}^{(1)}, \ldots, \widetilde{\boldsymbol{X}}^{(M)})|\boldsymbol{Y}, \boldsymbol{Z})\}$$
$$\sim = d_{TV}\{(\boldsymbol{X}|\boldsymbol{Y}, \boldsymbol{Z}), (\boldsymbol{X}_F|\boldsymbol{Y}, \boldsymbol{Z})\} = d_{TV}\{P(\cdot|\boldsymbol{Z}) P(\cdot|\boldsymbol{Z})\}.$$

将集合 $\mathcal{A}_\alpha$ 定义为

$$\mathcal{A}_\alpha := \Big\{(\boldsymbol{x}, \widetilde{\boldsymbol{x}}^{(1)}, \ldots, \widetilde{\boldsymbol{x}}^{(M)}) : \frac{1 + \sum_{m=1}^{M} I\{T(\widetilde{\boldsymbol{x}}^{(m)}, \boldsymbol{Y}, \boldsymbol{Z}) \ge T(\boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{Z})\}}{1+M} \le \alpha\Big\}.$$

然后，我们有

$$\begin{aligned} P(p \le \alpha|\boldsymbol{Y}, \boldsymbol{Z}) &= P\{(\boldsymbol{X}, \widetilde{\boldsymbol{X}}^{(1)}, \ldots, \widetilde{\boldsymbol{X}}^{(M)}) \in \mathcal{A}_\alpha|\boldsymbol{Y}, \boldsymbol{Z}\} \\ &\le d_{TV}\{((\boldsymbol{X}, \widetilde{\boldsymbol{X}}^{(1)}, \ldots, \widetilde{\boldsymbol{X}}^{(M)})|\boldsymbol{Y}, \boldsymbol{Z}), \\ &\qquad ((\widetilde{\boldsymbol{X}}_F, \widetilde{\boldsymbol{X}}^{(1)}, \ldots, \widetilde{\boldsymbol{X}}^{(M)})|\boldsymbol{Y}, \boldsymbol{Z})\} \\ &\quad + P\{(\widetilde{\boldsymbol{X}}_F, \widetilde{\boldsymbol{X}}^{(1)}, \ldots, \widetilde{X}^{(M)}) \in \mathcal{A}_\alpha|\boldsymbol{Y}, \boldsymbol{Z}\} \\ &= d_{TV}\{P(\cdot|\boldsymbol{Z}), \widehat{P}(\cdot|\boldsymbol{Z})\} \\ &\quad + P\{(\widetilde{\boldsymbol{X}}_F, \widetilde{\boldsymbol{X}}^{(1)}, \ldots, \widetilde{\boldsymbol{X}}^{(M)}) \in \mathcal{A}_\alpha|\boldsymbol{Y}, \boldsymbol{Z}\}. \end{aligned}$$

在以$\boldsymbol{Y}$和$\boldsymbol{Z}$为条件时，$\widetilde{\boldsymbol{X}}_F, \widetilde{\boldsymbol{X}}^{(1)}, \ldots, \widetilde{\boldsymbol{X}}^{(M)}$同分布，因此可交换，所以$P\{(\widetilde{\boldsymbol{X}}_F, \widetilde{\boldsymbol{X}}^{(1)}, \ldots, \widetilde{\boldsymbol{X}}^{(M)}) \in \mathcal{A}_\alpha|\boldsymbol{Y}, \boldsymbol{Z}\} \le \alpha$成立，并得到以下结果。

引理 3。假设 H_0：$X \perp\!\!\!\perp Y|Z$ 成立，对于任意给定的显著性水平 $\alpha \in (0,1)$，检验（5）的第一类错误满足

$$P(p \le \alpha|\boldsymbol{Y}, \boldsymbol{Z}) \le \alpha + d_{TV}\{P(\cdot|\boldsymbol{Z}), \widehat{P}(\cdot|\boldsymbol{Z})\}. \quad (9)$$

引理 3 的一个直接推论是，第一类错误率无条件成立如下：

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

此外，对于任何给定的检验统计量 $T(\cdots)$，我们都可以通过在式 (1) 中用 1-近邻样本 $\widetilde{\boldsymbol{X}}^{(m)}$ 替换 $\boldsymbol{X}^{(m)}$ 来计算 p 值。通过类似的论证，所得检验也满足式 (9)。

在 H_0, $I(\widetilde{X}^{(m)}; Y) \ge I(\widetilde{X}^{(m)}; Y|Z)$ 下。分别将（5）和（6）中的 p值记为 p 和 p^*。结合 $\widehat{I}(\widetilde{X}^{(m)}; Y|Z)$ 和 $\widehat{I}(\widetilde{X}^{(m)}; Y)$ 的一致性，我们得到以下主要结果。

定理 4。假设 H_0 成立，我们有

$$\begin{aligned} P(p^* \le \alpha|H_0) - \alpha &\le P(p \le \alpha|H_0) - \alpha \\ &\le E[d_{TV}\{P(\cdot|\boldsymbol{Z}), \widehat{P}(\cdot|\boldsymbol{Z})\}]. \end{aligned}$$

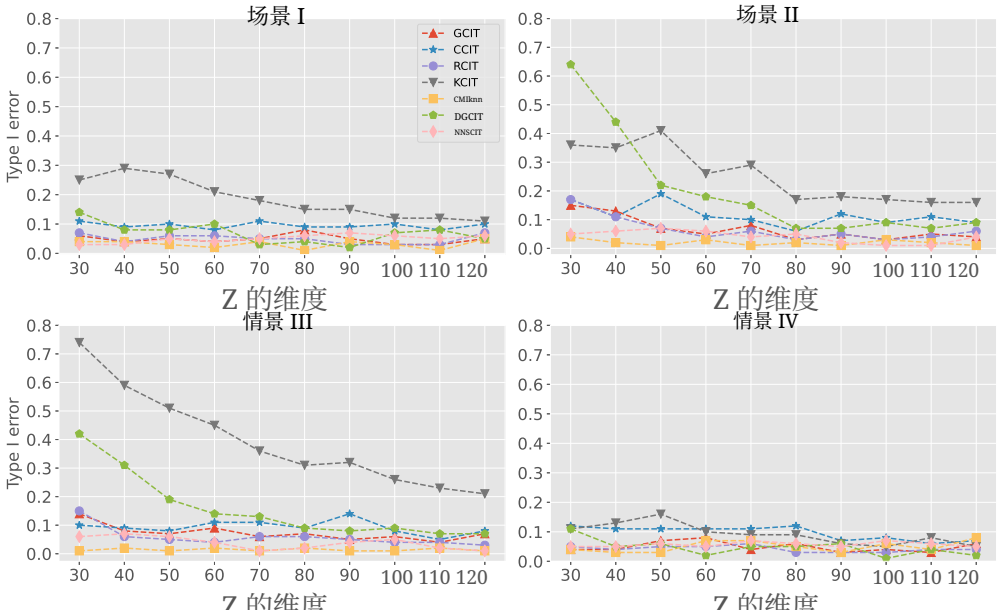


图 2： H_0 下各种检验的经验 I 类错误率。

d_Z	GCIT	CCIT	RCIT	KCIT	CMIknn		DGCIT	NNSCIT
5	0.11	0.36	0.53	0.72	0.03		0.50	0.01
10	0.48	0.22	0.82	0.84	0.06		0.86	0.05
15	0.87	0.21	0.85	0.93	0.15		0.91	0.05
20	0.93	0.25	0.90	0.98	0.05		0.97	0.07

表 1：示例 1 中各种检验的经验 I 类错误率。

定理4 有三个重要含义。首先，检验$\hat{}$(6) 超出期望水平 $\alpha \in$ (0,1) 的第一类错误由 $E\{d_{TV}(P(\cdot|\boldsymbol{Z})\ P(\cdot|\boldsymbol{Z}))$ }界定。其次，我们提出的方法在控制第一类错误方面优于 CRT (5)。第三，根据定理 2，我们得到

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

因此，我们的 NNSCIT 的额外第一类错误得以保证很小。

性能评估

在本节中，我们使用合成数据集考察 NNSCIT 在有限样本下的性能。我们将 NNSCIT 与 GCIT（Bellot 和 van der Schaar，2019）、基于分类器的条件独立性检验（CCIT）（Sen 等人，2017）、基于核的条件独立性检验（KCIT）（Zhang 等人，2011）、RCIT（Strobl、Zhang 和 Visweswaran，2019）、基于条件互信息的条件独立性检验（CMIknn）（Runge，2018）以及 DGCIT（Shi 等人，2021）进行比较。我们将一些额外的模拟研究和真实数据分析放在补充材料中。NNSCIT 的源代码可在 https://github.com/LeeShuai-kenwitch/NNSCIT 获取。

合成数据集上的性能

合成数据集采用与 Zhang 等人（2011年）和 Doran 等人类似的后非线性模型生成（2014年）以及 Bellot 和 van der Schaar（2019年）中的模型。具体而言，我们定义了在 H_0 和 H_1 条件下的（X，Y，Z）：

$$H_0 : X = f(A_f^T Z + \epsilon_f),\ Y = g(A_g^T Z + \epsilon_g),$$
$$H_1 : Y = h(A_h^T Z + bX) + \epsilon_h.$$

A_f 和 A_g 的元素从 $[0,1]$ 中随机且均匀地采样，然后归一化为单位 l_1 范数。 A_h 的元素从标准正态分布中采样，并将 b 设为2。噪声变量ϵ_f、 ϵ_g 和 ϵ_h分别独立地从均值为零、方差为 0.49 的正态分布中采样。显著性水平设为 $\alpha = 0.05$，样本量固定为 $n = 1000$。设置 $M = 500$ 和 $k = 3$。考虑以下四种情形：

场景I. 将 f、 g 和 h 设为恒等函数，从而在 H_1 下引入线性依赖关系、 $Z \sim N(0.7, 1)$ 和 $X \sim N(0, 1)$。

场景 II。将 f、 g 和 h 设置为场景 I 中的值，但使用拉普拉斯分布生成 Z。

情景 III。按照场景 I 设置 f、 g 和 h，但使用均匀分布 $[-2.5, 2.5]$ 生成 Z。

情景 IV。将 f、 g 和 h 设置为从 $\{x^2, x^3,$双曲正切函数$(x),$余弦函数$(x)\}$ 中随机采样。将 $Z \sim N$ 设置为 (0,1)，并在 H_1 下将 $X \sim N$ 设置为 $(0, 1)$。

我们将 Z 的维度设为 $d_Z = 30$、40、50、60、70、80、90、100、110和 120。图 2 和图 3 展示了类型

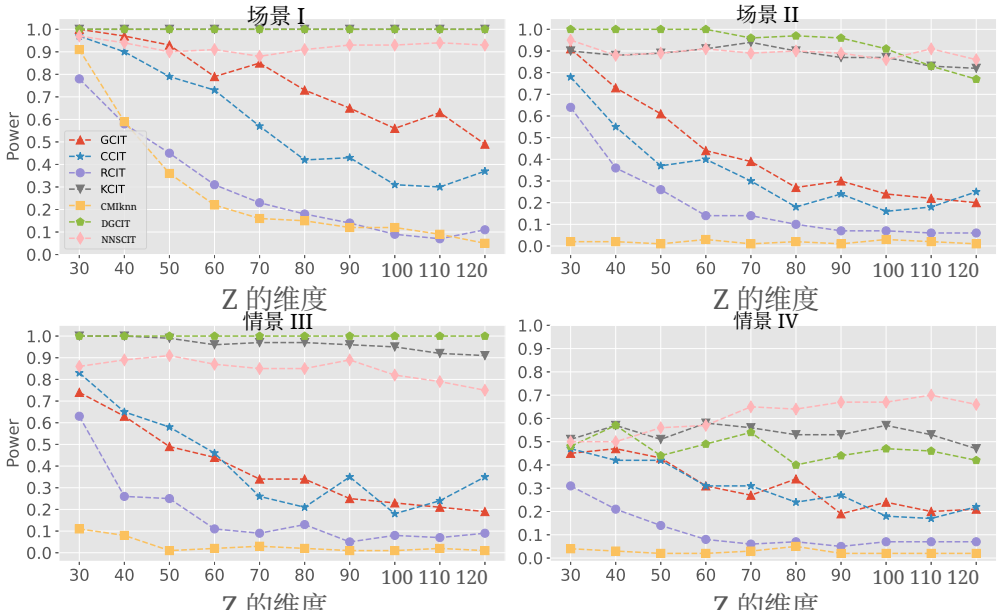


图 3： H_1 下各种检验的经验功效。

d_Z	GCIT	CCIT	RCIT	KCIT	CMIknn		DGCIT	NNSCIT
5	0.37	1	0.03	0.02	1		0.78	1
10	0.53	1	0.04	0.10	1		0.82	1
15	0.55	1	0.05	0.14	1		0.79	1
20	0.63	1	0.07	0.16	1		0.88	1

表2：示例2中各种检验的经验功效。

在 H_0 下的 I 类错误率，以及在 H_1下的检验功效，每种情况均基于 300 次数据重复模拟。$d_Z = 5$、10、15、20 和 25 的其他模拟结果见补充材料（图 1 和图 2）。

我们有以下观察结果。首先，我们的检验在 H_0下能够很好地控制第一类错误，同时在 H_1下具有较高的检验功效。其次，CMIknn在控制第一类错误方面表现良好，但在 H_1下，几乎在所有场景中都失去了检验功效。第三，尽管DGCIT和KCIT在 H_1下具有足够的检验功效，但在某些情况下它们的第一类错误率会膨胀，尤其是当 d_Z 小于30时。第四，GCIT、CCIT和RCIT在某些情况下无法控制第一类错误，尤其是当 d_Z 小于30时。此外，在 H_1下，GCIT、CCIT和RCIT在几乎所有场景中都损失了一定的检验功效。

补充材料中的图 4报告了在场景II下生成的数据进行单次 CIT 时，运行时间随 d_Z 变化的情况。其他场景表现相似。我们的NNSCIT 在计算上非常高效。相比之下，CCIT、CMIknn 和 DGCIT 耗时很长，在实际应用中难以实施。

两个示例的表现

如引言中所述，我们在以下两个示例中评估了我们方法的性能。The 数据生成机制的详细信息见补充材料。

示例 1. $X \to Z \to Y$。在这种情况下，H_0 成立，但 X 与 Y之间存在很强的依赖关系。表 1 报告了 I 类错误率。我们的 NNSCIT 对 I 类错误的控制效果很好，但 GCIT、CCIT、RCIT、KCIT 和 DGCIT 的表现不佳，因为它们的 I 类错误率非常高。

示例 2. $X \to Z \leftarrow Y$。在这种情况下，H_1 成立，但 X 与 Y 相互独立。表 2 报告了不同方法的检验功效。我们的方法检验功效高达 1。相比之下，RCIT 和 KCIT 的检验功效低于 0。2，GCIT 和 DGCIT 的检验功效也有所下降。

结论

本文提出了一种新颖且快速的 NNSCIT。我们使用 1近邻采样策略来近似条件分布 $X|Z$。与 GAN 相比，1近邻不仅具有计算效率，还在近似精度方面表现出优势。我们采用基于分类器的条件互信息（CCMI）估计量作为检验统计量，它能够很好地捕捉条件依赖特征。我们表明，NNSCIT 具有三个显著特点：能够良好地控制第一类错误率、在 H_1 下具有较高的检验功效，以及计算效率高。

致谢

Ziqi Chen 博士的工作部分得到了国家自然科学基金（NSFC）（12271167 和11871477）以及上海市自然科学基金（21ZR1418800）的支持。Christina Dan Wang 博士的工作部- 分得到了国家自然科学基金（NSFC）（12271363 和 11901395）的支持。

参考文献

Arjovsky, M.；Chintala, S.；和 Bottou, L. 2017。Wasserstein 生成对抗网络。载于 国际机器学习会议，214–223。PMLR。

Belghazi, M. I.; Baratin, A.; Rajeshwar, S.; Ozair, S.; Ben- gio,Y.; Courville,A.; and Hjelm, D. 2018. 互信息神经估计. In 国际机器学习会议,531–540. PMLR.

Bellot, A.；以及 van derSchaar, M. 2019。使用生成对抗网络进行条件独立性检验。载于神经信息处理系统进展，2199–2208。Berrett, T. B.；Wang,Y.；Barber, R. F.；以及Samworth,R. J. 2020。控制混杂因素时用于独立性检验的条件置换检验。英国皇家统计学会会刊：*B*辑（统计方法学），82(1)：175–197。Cai,Z.；Li,R.；以及Zhang, Y. 2022。具有因果发现应用的无分布条件独立性检验。机器学习研究期刊，23(85)：1–41。Candes,E.；Fan,Y.；Janson,L.；以及Lv, J. 2018。淘金：用于高维受控变量选择的 “model-X”knockoffs。英国皇家统计学会会刊：*B*辑（统计方法学），80(3)：551–577。Cover, T.；以及Hart, P. 1967。最近邻模式分类。*IEEE*信息论汇刊，13(1)：21–27。Cover, T. M.；以及 Thomas, J. A. 2012。信息论要素。JohnWiley & Sons。Devroye, L.；Gyorfi, L.；Krzyzak,A.；以及Lugosi, G. 1994。关于最近邻回归函数估计的一致强普适性。统计年刊，22(3)：1371–1385。Doran,G.；Muandet, K.；Zhang,K.；以及 Sch¨olkopf, B. 2014。基于置换的核条件独立性检验。载于人工智能中的不确定性，132–141。Citeseer。Fix, E.；以及 Hodges, J. L.1951。判别分析，非参数判别：一致性性质。技术报告4，美国空军航空医学学校。Fleuret, F. 2004。利用条件互信息进行快速二元特征选择。机器学习研究期刊，5(9)：1531–1555。Gao,W.；Oh,S.；以及Viswanath, P. 2016。突破带宽壁垒：几何自适应熵估计。载于神经信息处理系统进展，2460–2468。

Gao, W.; Oh, S.; and Viswanath, P. 2017. 揭开固定 k 近邻信息估计器的奥秘。载于*IEEE* 国际信息论研讨会，1267–1271。Genevay, A.; Peyr´e, G.; and Cuturi, M. 2018. 使用 Sinkhorn 散度学习生成模型。载于国际人工智能与统计会议，1608–1617。PMLR。Hlinka, J.; Hartman, D.; Vejmelka, M.; Runge, J.; Marwan, N.; Kurths, J.; and Palu˘s, M. 2013. 利用条件互信息推断有向气候网络的可靠性。*Entropy*，15(6)：2023–2045。Koller, D.; and Friedman, N. 2009. 概率图模型：原理与技术。麻省理工学院出版社。Lauritzen, S. L. 1996. 图模型，第17卷。克拉伦登出版社。Liang, K.-C.; and Wang, X. 2008. 使用条件互信息重建基因调控网络。*EURASIP* 生物信息学与系统生物学期刊，2008：1–14。Lopez-Paz, D.; and Oquab, M. 2017. 重新审视分类器双样本检验。载于国际学习表征会议。Mesner, O. C.; and Shalizi, C. R. 2020. 混合型、离散型和连续型数据的条件互信息估计。*IEEE* 信息论汇刊，67(1)：464–484。Mukherjee, S.; Asnani, H.; and Kannan, S. 2020. CCMI：基于分类器的条件互信息估计。载于人工智能中的不确定性，1083–1093。PMLR。Pearl, J. 2009. 因果关系。剑桥大学出版社。Runge, J. 2018. 基于条件互信息近邻估计器的条件独立性检验。载于国际人工智能与统计会议，938–947。PMLR。Sen, R.; Suresh, A. T.; Shanmugam, K.; Dimakis, A. G.; and Shakkottai, S. 2017. 模型驱动的条件独立性检验。载于神经信息处理系统进展，2955–2965。Shah, R. D.; and Peters, J. 2020. 条件独立性检验的难度及广义协方差测度。统计年刊，48(3)：1514–1538。Shi, C.; Xu, T.; Bergsma, W.; and Li, L. 2021. 用于条件独立性检验的双生成对抗网络。机器学习研究期刊，22(285)：1–32。Silverman, B. W. 2018. 统计与数据分析中的密度估计。劳特利奇出版社。Spirtes, P.; Glymour, C. N.; Scheines, R.; and Heckerman, D. 2000. 因果关系、预测与搜索。麻省理工学院出版社。Strobl, E. V.; Zhang, K.; and Visweswaran, S. 2019. 用于快速非参数因果发现的近似基于核的条件独立性检验。因果推断期刊，7(1)：1–24。Zhang, H.; Zhou, S.; Zhang, K.; and Guan, J. 2017. 使用基于回归的条件独立性进行因果发现

检验。载于人工智能促进协会人工智能会议论文集，1250–1256。Zhang, H.;Zhou, S.;Zhang, K.;and Guan,J. 2022. 基于残差相似性的条件独立性检验及其在因果发现中的应用。载于人工智能促进协会人工智能会议论文集，5942–5949。Zhang, K.; Peters, J.; Janzing, D.; and Sch¨olkopf, B. 2011. 基于核的条件独立性检验及其在因果发现中的应用。载于第二十七届人工智能不确定性会议论文集，804–813。