arXiv:2301.00967v1 [stat.ME] 3 Jan 2023

快速且准确的基于核的独立性检验及其在高维数据和函数型数据中的应用

Jin-Ting Zhang[a] 和 Tianming Zhu[b]

[a]新加坡国立大学统计与数据科学系，新加坡[b]南洋理工大学国立教育学院，新加坡

2023年1月4日

摘要

检验两个随机变量之间的依赖关系是统计学中的一个重要推断问题，因为许多统计程序都依赖于两个样本相互独立这一假设。为检验两个样本是否独立，研究者提出了一种所谓的基于希尔伯特–施密特独立性准则（HSIC）的检验。其零分布通过置换法或伽马近似进行逼近。本文提出了一种新的基于HSIC 的检验，并推导出其渐近零分布和备择分布。结果表明，所提出的检验具有根-n 一致性。本文采用三累积量匹配卡方近似来逼近检验统计量的零分布。通过选择适当的再生核，所提出的检验可应用于多种不同类型的数据，包括多元数据、高维数据和函数型数据。三项模拟研究和两个真实数据应用表明，在水平准确性、检验功效和计算成本方面，对于多元数据、高维数据和函数型数据，所提出的检验优于若干现有检验。

关键词 **RDS**：两样本独立性检验；希尔伯特-施密特独立性准则；三阶累积量匹配的 χ^2 近似 n.

1 引言

随着数据收集技术的发展，在各个领域中经常会遇到复杂的数据对象，例如高维数据或某些可分度量空间中的函数型数据。如今，在许多大数据应用中，我们通常关注于度量一对潜在的高维随机向量或函数型随机变量之间的关联程度。检验随机元素的独立性是统计学中的一个重要推断问题，并具有重要的应用价值。本研究的工作受到加拿大天气数据集的启发，该数据集可从 R 软件包fda.usc 中获取，且 Ramsay 和 Silverman (2005) 对其进行了详细讨论。该加拿大天气数据集已被研究

于多元函数型数据分析文献中；参见 Górecki 和 Smaga(2017) 以及Zhu 等人(2022。该数据集包含35 个加拿大气象站一年内的日平均温度曲线和日均降水曲线，这些曲线是通过对 1960 年至 1994年期间每年的日均温度曲线和日均降水曲线取平均得到的。我们感兴趣的是检验日平均温度曲线和日均降水曲线在统计上是否独立。这是一个针对函数型数据的两样本独立性检验问题。如果上述检验问题被拒绝，我们可以在推断程序中考虑这种依赖关系，从而提高效率。

从数学角度来说，两样本独立性检验问题可描述如下。设 x 和 y分别是定义在两个可分度量空间 $\mathcal{X}$ 和 $\mathcal{Y}$中的随机元素。假设我们有一个配对样本：

$$(x_i, y_i),\ i = 1, \ldots, n, \tag{1}$$

其中，每个 $(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}$ 均独立同分布，服从联合博雷尔概率测度P_{xy}。我们感兴趣的是检验以下假设：

$$H_0 : P_{xy} = P_x P_y, \quad \text{versus} \quad H_1 : P_{xy} \neq P_x P_y, \tag{2}$$

其中， P_x 和 P_y 分别是 x 和 y的边缘概率测度。

已有一些经典的依赖度量，例如 Spearman 的 ρ 和 Kendall 的 τ，并得到了广泛应用。然而，这些度量通常只用于捕捉特定形式的依赖关系（例如线性依赖或单调依赖），无法检测随机变量之间的所有依赖模式。随着复杂数据对象日益普遍，人们需要能够捕捉更复杂依赖模式的依赖度量，包括高维数据集或函数型数据集之间的依赖模式。

希尔伯特–施密特独立性准则（希尔伯特-施密特独立性准则）由 Gretton 等人（2005a,b）提出并加以研究，是最成功的非参数依赖度量之一。它利用再生核希尔伯特空间（再生核希尔伯特空间，RKHS）中概率测度的核嵌入之间的距离（Gretton 等人 2007；Smola 等人 2007；Zhang 等人 2011），可用于衡量单变量或多变量随机变量之间的依赖关系，也可用于衡量取值属于更复杂结构（如高维数据或函数型数据）的随机变量之间的依赖关系。Gretton 等人（2007）通过采用 HSIC，提出了一种新颖的检验，其检验统计量是利用 V 统计量得到的 HSIC 经验估计。作者用双参数伽马分布近似检验统计量的零分布。由此得到的基于伽马近似的检验的计算成本为 $\mathcal{O}(n^2)$，其中 n是样本量。然而，表1和表3中的模拟结果表明，基于伽马近似的检验仅在维度 p 较小时表现良好；当维度p 较大时，该检验非常保守，甚至完全失效。Gretton 等人（2007）还提出了一种置换检验，该检验通常具有良好的尺寸控制表现，但需要进行 $\mathcal{O}(mn^2)$ 次运算，其中 m 是置换次数，这意味着其耗时约为基于伽马近似的检验的 m 倍。第3.1节的表2部分证实了这一点，该表显示置换检验的耗时约为

比基于伽马近似的检验高出 100 至 1000 倍。为解决这一问题，Zhang 等人(2018) 提出了三种 HSIC 快速估计量，以加快基于希尔伯特-施密特独立性准则的检验中的计算。然而，使用这些快速估计量降低计算复杂度并非毫无代价。根据Zhang 等人(2018) 的模拟结果，当样本量较小时，检验功效会有所损失。因此，大规模示例中计算成本的大幅降低，抵消了对更大样本量的需求。

近年来，函数型数据分析已成为统计学中的一个重要领域。大多数研究在未进行任何检验的情况下，假设随机曲线彼此独立。为解决这一问题，人们开发了一些检测随机曲线之间依赖关系的方法。这些独立性检验大多基于相关性度量，包括经典的皮尔逊相关（Pearson（1895））、动态相关（Dubin 和 Müller（2005））和全局时间相关（Zhou 等人（2018））等。然而，由于零相关通常并不意味着独立性，这些函数型相关可能不足以用于独立性检验（Miao 等人（2022））。Kosorok (2009) 将 Székely 等人提出的距离协方差 (2007) 应用于前几个函数主成分得分，这些得分累计解释了随机函数变异的 95 %。遗憾的是，正如 Shen 等人 (2019) 所讨论的，以及表 6 中的模拟结果所示，在检验两个随机函数之间的依赖关系时，相关性检验和基于距离协方差的检验对单调关系具有较高的检验功效，但对非单调依赖关系的检验功效较低。

本文提出了一种新的基于 HSIC 的检验，适用于多元、高维和函数型数据，且计算速度非常快。据我们所知，能够良好适用于多元、高维和函数型数据的检验为数不多。本研究的主要贡献如下。首先，我们提出了一个无偏且具有根n 一致性的估计量，用于估计所提出检验统计量中采用的中心化再生核。这为所提出的新检验提供了良好基础，使其尺寸控制显著优于Gretton 等人(2007)提出的基于伽马近似的检验。其次，在一些正则条件下，我们证明了原假设下所提出的检验统计量具有卡方型混合极限。第三，我们推导出所提出检验统计量的前三个累积量（均值、方差和三阶中心矩）。这使我们能够采用χ^2-近似参数由数据一致估计的Zhang(2005)三累积量匹配 χ^2-近似，准确近似卡方型混合分布。三累积量匹配近似无需置换，并能显著降低计算成本。这保证了所提出的新检验计算速度非常快，且具有良好的尺寸控制。第四，我们推导出所提出的新检验在局部备择假设下的渐近功效，并证明其具有根n 一致性。据我们所知，文献中尚未对此进行研究。最后，通过三项模拟研究和两个真实数据示例，我们表明，就尺寸控制、检验功效和计算成本而言，我们的新检验表现良好，并且在针对多元、高维和函数型数据的独立性检验中优于若干现有检验。

本文其余部分安排如下。主要结果见第2节。模拟研究和真实数据应用分别见第3 节和第4节。总结见第5节。主要结果的技术性证明概述于附录中。

2 主要结果

2.1 检验统计量

设 $K(\cdot,\cdot): \mathcal{X}\times\mathcal{X}\to\mathcal{R}$ 和 $L(\cdot,\cdot): \mathcal{Y}\times\mathcal{Y}\to\mathcal{R}$ 是两个连续、具有正特征的再生核。设 K 和 L 分别生成的两个再生核希尔伯特空间（再生核希尔伯特空间）为 $\mathcal{F}$ 和 $\mathcal{G}$，其内积分别为 $\langle,\rangle_{\mathcal{F}}$ 和 $\langle,\rangle_{\mathcal{G}}$。设 $\phi(x) = K(x,\cdot)$ 和 $\psi(y) = L(y,\cdot)$ 表示它们相应的典范特征映射，因此有以下核技巧：

$$K(x,x') = \langle\phi(x),\phi(x')\rangle_{\mathcal{F}}, \text{ and } L(y,y') = \langle\psi(y),\psi(y')\rangle_{\mathcal{G}},$$

其中 (x',y') 是 (x,y)的一个独立副本。因此 $\phi(\mathcal{X}) \subset \mathcal{F}$和 $\psi(\mathcal{Y}) \subset \mathcal{G}$。根据Fukumizu 等人 (2004)，交叉协方差算子 $\mathcal{C}_{xy}: \mathcal{G}\to\mathcal{F}$ 定义为对所有 $f\in\mathcal{F}$ 和$g\in\mathcal{G}$

$$\langle f,\mathcal{C}_{xy}g\rangle_{\mathcal{F}} = \mathrm{Cov}[f(x),g(y)]. \tag{3}$$

令 $\mu_x = \mathrm{E}_x[\phi(x)]$ 和 $\mu_y = \mathrm{E}_y[\psi(y)]$ 分别为概率测度 P_x 和 P_y的均值嵌入。于是，交叉协方差算子本身可写为

$$\mathcal{C}_{xy} = \mathrm{E}_{xy}\{[\phi(x)-\mu_x]\otimes[\psi(y)-\mu_y]\},$$

其中 $\otimes$ 表示张量积。

根据 Gretton 等人（2005b，定理4），当且仅当对所有连续有界函数 $f(x)$ 和 $g(y)$，$\mathrm{Cov}[f(x)\ g(y)] = 0$ 时，x和 y相互独立。于是，表达式（3）表明，当且仅当 $\mathcal{C}_{xy} = 0$时，x 和 y相互独立。也就是说，当且仅当它们的核嵌入 $\phi(x)$ 和 $\psi(y)$ 不相关时，x 和 y 相互独立。因此，使用原始独立同分布样本（1）检验（2），等价于检验以下假设：

$$H_0: \mathcal{C}_{xy} = 0, \text{ versus } H_1: \mathcal{C}_{xy}\neq 0. \tag{4}$$

Gretton 等人 (2007) 提出使用以下 x 与 y 的平方希尔伯特–施密特范数来度量二者之间的依赖关系 $\mathcal{C}_{xy}$：

$$\begin{aligned}\|\mathcal{C}_{xy}\|^2_{\mathcal{F}\otimes\mathcal{G}} &= \mathrm{E}_{xyx'y'}\left[\langle\phi(x)-\mu_x,\phi(x')-\mu_{x'}\rangle_{\mathcal{F}}\langle\psi(y)-\mu_y,\psi(y')-\mu_{y'}\rangle_{\mathcal{G}}\right]\\ &= \mathrm{E}_{xyx'y'}[\tilde{K}(x,x')\tilde{L}(y,y')],\end{aligned} \tag{5}$$

其中，$\tilde{K}(x,x')$ 和 $\tilde{L}(y,y')$ 分别表示 $K(x,x')$ 和 $L(y,y')$ 的中心化版本，而 (x',y') 是 (x,y) 的一个独立副本。注意，对于核 $K(\cdot,\cdot)$，我们有

$$\begin{aligned}\tilde{K}(x,x') &= \langle\phi(x)-\mu_x,\phi(x')-\mu_{x'}\rangle_{\mathcal{F}} \\ &= K(x,x')-\mathrm{E}_{z'}[K(x,z')]-\mathrm{E}_z[K(z,x')]+\mathrm{E}_{z,z'}[K(z,z')],\end{aligned} \tag{6}$$

其中，z 和 z' 分别是 x 和 x'的独立副本。还要注意，我们有以下有用性质：当 $x'=x$时，有

$$\mathrm{E}[\tilde{K}(x,x)]=\langle\phi(x)-\mu_x,\phi(x)-\mu_x\rangle_{\mathcal{F}}=\mathrm{E}\,\|\phi(x)-\mu_x\|_{\mathcal{F}}^2\geq 0,$$

当 x 和 x' 相互独立时，有

$$\mathrm{E}_x[\tilde{K}(x,x')]=\mathrm{E}_{x'}[\tilde{K}(x,x')]=\mathrm{E}_{x,x'}[\tilde{K}(x,x')]=0.$$

上述性质在将 K 和 x 分别替换为 L 和 y后仍然成立。利用（5），为了检验（4），我们可以构造如下检验统计量

$$T_n=n^{-1}\sum_{i=1}^n\sum_{j=1}^n\tilde{K}^*(x_i,x_j)\tilde{L}^*(y_i,y_j),$$

其中，$\tilde{K}^*(x_i,x_j)$ 和 $\tilde{L}^*(y_i,y_j)$ 分别是 $\tilde{K}(x_i,x_j)$ 和 $\tilde{L}(y_i,y_j)$ 的无偏估计量，其定义如下

$$\begin{aligned}\tilde{K}^*(x_i,x_j)&=K(x_i,x_j)-\sum_{v\neq i}\frac{K(x_i,x_v)}{n-1}-\sum_{u\neq j}\frac{K(x_u,x_j)}{n-1}+\sum_{u\neq v}\frac{K(x_u,x_v)}{n(n-1)},\ \text{and}\\ \tilde{L}^*(y_i,y_j)&=L(y_i,y_j)-\sum_{v\neq i}\frac{L(y_i,y_v)}{n-1}-\sum_{u\neq j}\frac{L(y_u,y_j)}{n-1}+\sum_{u\neq v}\frac{L(y_u,y_v)}{n(n-1)}.\end{aligned} \tag{7}$$

为简便起见，令 $\tilde{\mathrm{K}}=(\tilde{K}(x_i,x_j))：n\times n$和 $\tilde{\mathrm{L}}=(\tilde{L}(y_i,y_j))：n\times n$分别表示两个核 $\tilde{K}(\cdot,\cdot)$ 和 $\tilde{L}(\cdot,\cdot)$ 的格拉姆矩阵。类似地，令 $\tilde{\mathrm{K}}^*=(\tilde{K}^*(x_i,x_j))：n\times n$和 $\tilde{\mathrm{L}}^*=(\tilde{L}^*(y_i,y_j))：n\times n$。那么有

$$T_n=n^{-1}\,\mathrm{tr}(\tilde{\boldsymbol{K}}^*\tilde{\boldsymbol{L}}^*), \tag{8}$$

其中，$\mathrm{tr}(\boldsymbol{A})$ 表示方阵 $\boldsymbol{A}$ 的迹，即 $\boldsymbol{A}$ 的对角元素之和。注意，T_n 可以通过 $\mathcal{O}(n^2)$ 次运算轻松计算。

注释 1。为了检验（4），Gretton 等人（2007）提出了以下检验统计量：

$$T_{n,G}=n^{-1}\,\mathrm{tr}(\boldsymbol{HKHL})=n^{-1}\,\mathrm{tr}(\boldsymbol{K}_G\boldsymbol{L}_G)=n^{-1}\sum_{i=1}^n\sum_{j=1}^n K_G(x_i,x_j)L_G(y_i,y_j), \tag{9}$$

其中 $\mathrm{H}=\mathrm{I}_n-\mathrm{J}_n/n$，$\mathrm{I}_n$和 J_n分别为 $n\times n$单位矩阵和 $n\times n$全一矩阵；$\mathrm{K}_G=(K_G(x_i,x_j))：n\times n$以及 $\mathrm{L}_G=(L_G(y_i,y_j))：n\times n$，其中 $K_G(x_i,x_j)$ 和 $L_G(y_i,y_j)$ 是 $\tilde{K}(x_i,x_j)$ 和 $\tilde{L}(y_i,y_j)$ 的无偏估计量，定义如下

$$\begin{aligned}K_G(x_i,x_j)&=K(x_i,x_j)-\sum_{u=1}^n\frac{K(x_i,x_u)}{n}-\sum_{v=1}^n\frac{K(x_v,x_j)}{n}+\sum_{v=1}^n\sum_{u=1}^n\frac{K(x_v,x_u)}{n^2},\ \text{and}\\ L_G(y_i,y_j)&=L(y_i,y_j)-\sum_{u=1}^n\frac{L(y_i,y_u)}{n}-\sum_{v=1}^n\frac{L(y_v,y_j)}{n}+\sum_{v=1}^n\sum_{u=1}^n\frac{L(y_v,y_u)}{n^2}.\end{aligned} \tag{10}$$

值得强调的是，我们的检验统计量 T_n (8) 与 Gretton 等人(2007) 的检验统计量 $T_{n,\ G}$ (9) 之间的差异，主要源于无偏估计量 (7) 与有偏估计量 (10) 在估计 $K(x_i,\ x_j)$ 和 $L(y_i,\ y_j)$时的差异。

2.2 渐近零分布

假设

$$|K(x,x')| < B_K < \infty,\ \text{ and }\ |L(y,y')| < B_L < \infty \text{ for all } x,x' \in \mathcal{X} \text{ and } y,y' \in \mathcal{Y}, \tag{11}$$

其中 B_K和 B_L是两个常数。条件 (11) 保证 $|K(x,x')| \leq 4B_K$, $|L(y,y')| \leq 4B_L$, $\mathrm{E}[K(x,x)] < \infty$，以及 $\mathrm{E}[L(y,y)] < \infty$。由柯西–施瓦茨不等式可知，$K(x,x')$ 和 $L(y,y')$ 均平方可积，即 $\mathrm{E}[\tilde{K}^2(x,x')] < \infty$ 且 $\mathrm{E}[\tilde{L}^2(y,y')] < \infty$。因此，$\tilde{K}(x,x')$ 和 $\tilde{L}(y,y')$ 具有如下 Mercer 展开

$$\tilde{K}(x,x') = \sum_{r=1}^{\infty} \lambda_r \phi_r(x)\phi_r(x'),\ \text{ and }\ \tilde{L}(y,y') = \sum_{r=1}^{\infty} \rho_r \psi_r(y)\psi_r(y'), \tag{12}$$

其中 λ_1，λ_2，.....和 ρ_1，ρ_2，.....分别是 $K(x,\ x')$ 和 $L(y,\ y')$ 的特征值；此外，稍微滥用一下记号，$\phi_1(x)$ $\phi_2(x)$.....和 $\psi_1(y)$ $\psi_2(y)$.....分别是 $K(x,\ x')$ 和 $L(y,\ y')$ 的相关正交归一特征元素，其意义如下

$$\begin{aligned} \textstyle\int_{\mathcal{X}} \tilde{K}(x,x')\phi_r(x)P_x(dx) &= \lambda_r\phi_r(x'), \qquad \textstyle\int_{\mathcal{X}} \phi_r(x)\phi_s(x)P_x(dx) = \delta_{rs},\ r,s = 1,2,\ldots, \\ \textstyle\int_{\mathcal{Y}} \tilde{L}(y,y')\psi_r(y)P_y(dy) &= \rho_r\psi_r(y'), \qquad \textstyle\int_{\mathcal{Y}} \psi_r(y)\psi_s(y)P_y(dy) = \delta_{rs},\ r,s = 1,2,\ldots, \end{aligned}$$

其中，当 $r = s$ 时 $\delta_{rs} = 1$，否则为 0。为推导 T_n 的渐近零分布，我们需要以下引理，其证明见附录。

引理 1. 在条件 *(11)* 下，当 $n \to \infty$ 时，我们有

$$\tilde{K}^*(x_i,x_j) = \tilde{K}(x_i,x_j) + \mathcal{O}(n^{-1/2}) \textit{ uniformly for all } x_i,x_j.$$

引理 1 给出了 $K^*(x_i,x_j)$ 到 $K(x_i,x_j)$的一致收敛速率。类似地，我们也有$\tilde{L}^*(y_i,y_j) = \tilde{L}(y_i,y_j) + \mathcal{O}(n^{-1/2})$ 对所有 y_i,y_j均一致成立。因此，我们可以写为

$$T_n = n^{-1}\,\mathrm{tr}(\tilde{\boldsymbol{K}}\tilde{\boldsymbol{L}}) + \mathcal{O}(n^{-1/2}) \equiv \tilde{T}_n + \mathcal{O}(n^{-1/2}), \tag{13}$$

也就是说，对于较大的 n，T_n 和 T_n 具有相同的分布。因此，研究 T_n 的渐近零分布等价于研究 T_n 的分布。

定理 1. 在条件 *(11)* 和原假设下，当 $n \to \infty$ 时，我们有 $\tilde{T}_n \xrightarrow{\mathcal{L}} \tilde{T}$ 且 $T_n \xrightarrow{\mathcal{L}} \tilde{T}$，其中

$$\tilde{T} \overset{d}{=} \sum_{r=1}^{\infty}\sum_{s=1}^{\infty} \lambda_r\rho_s A_{rs},\ A_{rs} \overset{i.i.d.}{\sim} \chi_1^2,$$

其中，λ_1、λ_2、.....和 ρ_1、ρ_2、.....分别是 $K(x,x')$ 和 $L(y,y')$的特征值。

注 2。定理 1 与 Gretton 等人（2007）的定理 2 相对应，该文作者将$T_{n,G}/n$ （见注 1）视为一个四阶核的 V 统计量；然而，事实上我们可以证明，$T_{n,G}/n = \tilde{T}_n/n + \mathcal{O}(n^{-3/2})$ 中 $\tilde{T}_n/n$ 仅是一个二阶核的 V 统计量 [，详情见式（13）]。定理 1 与 Zhang 等人（2018）的定理 1 相同，但我们的证明比后者的证明简单得多。

2.3 零分布近似

如引言部分所述，Gretton 等人（2007）通过置换和双参数伽马分布近似了$T_{n、G}$ （9）的零分布，分别得到置换检验和基于伽马近似的检验。本小节中，由于 $\tilde{T}$ 是一个系数未知的 χ^2型混合分布，我们采用三累积量（3-c）匹配的 T_n 近似 χ^2（Zhang（2005）、Zhang（2013））， 近似 χ^2的零分布，从而得到一种基于三累积量（3-c）匹配近似的新检验。

备注3。在计算成本方面，置换检验非常耗时，其成本为$\mathcal{O}$（mn^2），其中 m 是置换次数， n 是样本量；而基于伽马近似的检验计算速度很快，成本为 $\mathcal{O}$（n^2）。然而，在尺寸控制方面，置换检验通常表现良好，但基于伽马近似的检验仅在低维数据上表现良好；对于高维数据，该检验过于保守，完全无法发挥作用，这一点可由第3.1节表1和表3中的模拟结果看出。

所提出的新检验的核心思想是利用以下形式的随机变量的零分布来近似 T_n 的零分布

$$R \overset{d}{=} \beta_0 + \beta_1 \chi^2_d.$$

参数 β_0， β_1 和 d 通过匹配 T_n 和 R 的前三阶累积量来确定。为此，我们推导出 T_n 的前三阶累积量，如下述定理所示，其证明见附录。

定理 2. 在条件（*11*）和原假设下， T_n 的前三阶累积量为

$$\begin{aligned} &\mathrm{E}(T_n) = M_1 N_1 + \mathcal{O}(n^{-1/2}), \quad \mathrm{Var}(T_n) = 2M_2 N_2 + \mathcal{O}(n^{-1/2}), \quad \textit{and} \\ &\mathrm{E}[T_n - \mathrm{E}(T_n)]^3 = 8M_3 N_3 + \mathcal{O}(n^{-1/2}), \end{aligned} \tag{14}$$

其中， x、 x'、 x'' 独立同分布$\sim$ P_x，且 y、 y'、 y'' 独立同分布$\sim$ P_y,

$$\begin{aligned} &M_1 = \mathrm{E}[\tilde{K}(x,x)], \quad M_2 = \mathrm{E}[\tilde{K}^2(x,x')], \quad M_3 = \mathrm{E}[\tilde{K}(x,x')\tilde{K}(x',x'')\tilde{K}(x'',x)], \\ &N_1 = \mathrm{E}[\tilde{L}(y,y)], \quad N_2 = \mathrm{E}[\tilde{L}^2(y,y')], \quad N_3 = \mathrm{E}[\tilde{L}(y,y')\tilde{L}(y',y'')\tilde{L}(y'',y)]. \end{aligned} \tag{15}$$

的前三个累积量 R可表示为 $\beta_0 + \beta_1 d$、 $2\beta_1^2 d$和 $8\beta_1^3 d$；而 T_n <style id='16} 的前三个累积量见式（14）。令 T_n 与 R 的前三个累积量相等，并忽略高阶项，可得

$$\beta_0 = M_1 N_1 - \frac{(M_2 N_2)^2}{M_3 N_3}, \quad \beta_1 = \frac{M_3 N_3}{M_2 N_2}, \quad \text{and} \quad d = \frac{(M_2 N_2)^3}{(M_3 N_3)^2}. \tag{16}$$

此外， T_n 的偏度也可近似表示为

$$\frac{\mathrm{E}[T_n-\mathrm{E}(T_n)]^3}{\mathrm{Var}^{3/2}(T_n)}=\frac{8M_3N_3}{(2M_2N_2)^{3/2}}=\sqrt{8/d}.$$

因此，随着 d 增大， T_n 的偏度将变小。

$\hat{}$ $\hat{}$ $\hat{}$ $\hat{}\hat{}\hat{}$令 M_1， M_2， M_3和 N_1， N_2， N_3分别为 M_1， M_2， M_3和 N_1， N_2， N_3的相合估计量。将它们代入（16），即可得到 β_0， β_1 和 d 的相合估计量:

$$\hat{\beta}_0=\hat{M}_1\hat{N}_1-\frac{(\hat{M}_2\hat{N}_2)^2}{\hat{M}_3\hat{N}_3},\ \hat{\beta}_1=\frac{\hat{M}_3\hat{N}_3}{\hat{M}_2\hat{N}_2},\ \text{and}\ \hat{d}=\frac{(\hat{M}_2\hat{N}_2)^3}{(\hat{M}_3\hat{N}_3)^2}. \tag{17}$$

那么，对于任意名义显著性水平 $\alpha>0$，令 $\chi^2_d(\alpha)$ 表示 100α 的上侧 χ^2_d百分位数。然后，利用（17），采用 3-c 匹配 χ^2近似的所提出的新检验可通过使用近似临界值 $\hat{}\ \beta_0+\hat{}\beta_1\chi^2_{\hat{d}}(\alpha)$ 或近似 p值 $P\{\chi^2_{\hat{d}}\geq(T_n-\hat{}\beta_0)/\hat{}\beta_1\}$。

在条件（11）下，根据引理 1，对于所有 x_i,x_j，我们一致地有 $\tilde{K}^*(x_i,x_j)=\tilde{K}(x_i,x_j)+\mathcal{O}(n^{-1/2})$; 对于所有 y_i,y_j，也一致地有 $\tilde{L}^*(y_i,y_j)=\tilde{L}(y_i,y_j)+\mathcal{O}(n^{-1/2})$。然后根据（15）， M_1,M_2,M_3 和 N_1, N_2, N_3 的自然估计量为

$$\begin{aligned}\hat{M}_1&=\tfrac{1}{n}\sum_{i=1}^{n}\tilde{K}^*(x_i,x_i),\quad \hat{M}_2=\tfrac{2}{n(n-1)}\sum_{1\leq i<j\leq n}[\tilde{K}^*(x_i,x_j)]^2,\\ \hat{M}_3&=\tfrac{6}{n(n-1)(n-2)}\sum_{1\leq i<j<k\leq n}[\tilde{K}^*(x_i,x_j)\tilde{K}^*(x_j,x_k)\tilde{K}^*(x_k,x_i)],\end{aligned} \tag{18}$$

and

$$\begin{aligned}\hat{N}_1&=\tfrac{1}{n}\sum_{i=1}^{n}\tilde{L}^*(y_i,y_i),\quad \hat{N}_2=\tfrac{2}{n(n-1)}\sum_{1\leq i<j\leq n}[\tilde{L}^*(y_i,y_j)]^2,\\ \hat{N}_3&=\tfrac{6}{n(n-1)(n-2)}\sum_{1\leq i<j<k\leq n}[\tilde{L}^*(y_i,y_j)\tilde{L}^*(y_j,y_k)\tilde{L}^*(y_k,y_i)],\end{aligned}$$

其中 $\tilde{K}^*(x_i,x_j)$ 和 $\tilde{L}^*(y_i,y_j)$ 定义于式（7）中。为加快计算，我们可以写作 $\hat{}\ M_1=\mathrm{tr}(\ \tilde{}\boldsymbol{K}^*)/n$,

$$\begin{aligned}\hat{M}_2&=[n(n-1)]^{-1}\left[\mathrm{tr}(\tilde{\boldsymbol{K}}^{*2})-\mathrm{tr}(\tilde{\boldsymbol{K}}^{*}o\tilde{\boldsymbol{K}}^{*})\right],\ \text{and}\\ \hat{M}_3&=[n(n-1)(n-2)]^{-1}\left\{\mathrm{tr}(\tilde{\boldsymbol{K}}^{*3})-3\,\mathrm{tr}[\mathrm{diag}(\tilde{\boldsymbol{K}}^{*})\tilde{\boldsymbol{K}}^{*2}]+2\,\mathrm{tr}(\tilde{\boldsymbol{K}}^{*}o\tilde{\boldsymbol{K}}^{*}o\tilde{\boldsymbol{K}}^{*})\right\},\end{aligned} \tag{19}$$

其中，AoB $=(a_{ij}b_{ij})$ 表示矩阵 A $=(a_{ij})$ 和 B $=(b_{ij})$ 的点积，而 diag(A) 表示由 A 的对角元素构成的对角矩阵。式（19）的证明见附录。类似地，我们有 $\hat{}\ N_1=\mathrm{tr}(\tilde{}\mathrm{L}^*)/n$, $\hat{}N_2=[n(n-1)]^{-1}[\mathrm{tr}(\tilde{}\mathrm{L}^{*2})-\mathrm{tr}(\tilde{}\mathrm{L}^{*}o\tilde{}\mathrm{L}^{*})]$, 以及

$$\hat{N}_3=[n(n-1)(n-2)]^{-1}\left\{\mathrm{tr}(\tilde{\boldsymbol{L}}^{*3})-3\,\mathrm{tr}[\mathrm{diag}(\tilde{\boldsymbol{L}}^{*})\tilde{\boldsymbol{L}}^{*2}]+2\,\mathrm{tr}(\tilde{\boldsymbol{L}}^{*}o\tilde{\boldsymbol{L}}^{*}o\tilde{\boldsymbol{L}}^{*})\right\}.$$

定理 3. 在条件（11）下，当 $n\to\infty$ 时，有 $\hat{}M_\ell\xrightarrow{p}M_\ell$， $\ell=1,2,3$ 以及 $\hat{}N_\ell\xrightarrow{p}N_\ell$， $\ell=1,2,3$。由此可得，当 $n\to\infty$ 时，有 $\hat{}\beta_0\xrightarrow{p}\beta_0$， $\hat{}\beta_1\xrightarrow{p}\beta_1$，以及 $\hat{}d\xrightarrow{p}d$.

备注 4. 由于基于伽马近似的检验匹配了 T_n 的均值和方差，G 而所提出的新检验匹配了 T_n 的均值、方差和三阶中心矩，因此在尺寸控制方面，预计所提出的新检验应显著优于基于伽马近似的检验。表 1 和 3 中的模拟结果

第 3.1 节。此外，就计算成本而言，所提出的新检验的计算成本为 $\mathcal{O}(n^3)$，当置换次数大于样本量时，其耗时远少于置换检验，但略多于基于伽马近似的检验。第 3.1 节的表 2对此提供了部分佐证。

2.4 渐近功效

在本小节中，我们研究所提出的检验在以下局部备择假设下的渐近功效:

$$H_{1n}: \text{HSIC} = \|\mathcal{C}_{uv}\|_{\mathcal{F}\otimes\mathcal{G}} = n^{-(1/2-\Delta)}h, \tag{20}$$

其中 $0 < \Delta < 1/2$ 且 h 是一个正数。随着样本量 n 趋于无穷大，上述局部备择假设将趋近于原假设。因此，通常难以检测到它。如果一个检验能够以趋近于n的概率检测出局部备择假设（20），则通常称其为根1 一致的， n 趋于无穷大时亦然。根n 一致检验通常更受青睐。

定理4. 在条件（*11*）和局部备择假设（*20*）下，当 $n \to \infty$时，我们有

$$\sqrt{n}\,(T_n/n - HSIC) \xrightarrow{\mathcal{L}} N(0, 4\sigma^2),$$

其中，(x', y') 是 (x, y) 的独立副本，$\sigma^2 = \text{Var}\left\{\text{E}[\tilde{K}\,(x, x')\,\tilde{L}\,(y, y')\mid (x', y')\,]\right\}$。因此，对于任意显著性水平 α，所提出检验 T_n 的渐近功效为

$$Pr\left[T_n \geq \hat{\beta}_0 + \hat{\beta}_1\chi^2_{\tilde{d}}(\alpha)\right] = \Phi\left[n^{\Delta}h/(2\sigma)\right][1 + o(1)],$$

其趋于1，当 $n \to \infty$ 其中 $\Phi(\cdot)$ 表示 $N(0,\ 1)$的累积分布函数。

定理 4的证明见附录。定理 4表明，所提出的新检验 T_n是根n 一致的。

注 5。 定理4的第一个结果与 Gretton 等人 (2007) 的定理 1 类似，其中作者将 $T_{n,G}/n$ 视为一个 4 阶核的 V 统计量，但实际上我们有 $T_{n,G}/n = \tilde{T}_n/n + O_p(n^{-3/2})$~，其中 T_n/n 仅是一个 2 阶核的 V 统计量 [，详见（13）]。Gretton 等人Gretton et al. (2007) 的定理 1 可能存在问题。

3 模拟研究

在本节中，我们开展三项模拟研究，即模拟 1、2 和 3，分别针对多元数据、高维数据和函数型数据的两样本独立性检验问题，将所提出的新检验 NEW 与若干现有竞争方法进行比较。我们将检验的经验尺寸或检验功效计算为 10，000 次模拟运行中拒绝次数所占的比例。在本节中，我们将名义水平 α 设为 5 %。

在下文所述的三项模拟研究中，为简便起见，我们选择核 $K(\cdot,\ \cdot)$ 为以下高斯径向基函数（RBF）核：

$$K(x,x') = \exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right),$$

其中 σ^2 是所谓的核宽度。对于模拟1和模拟2中的多元和高维数据，$\|x\|$表示向量的通常 L^2-范数 x；对于模拟3中的函数型数据，它表示函数 L^2的通常 $x(t)\ t \in \mathcal{T}$ 范数，其定义为 $\|x\| = \left[\int_{\mathcal{T}} x^2(t)dt\right]^{1/2}$，并通过使用梯形法则近似积分来计算。显然，上述高斯径向基函数核的上界为 1，因此条件（11）始终成立。核宽度 σ^2通过采用 Zhang 等人提出的数据自适应高斯核宽度选择方法进行选择，参见 Zhang 等人 (2022，第 2.6 节）。对于核 $L(\cdot,\cdot)$，方法类似。

3.1 模拟1

在这项模拟研究中，我们展示了 NEW 检验在多元数据上的表现，并将其与基于伽马近似的检验以及 Gretton 等人(2007)提出并研究的置换检验进行比较，二者分别记为 HSICg 和 HSICp。HSICg 检验通过 R 软件包dHSIC (Pfister 和 Peters 2017)实现，HSICp 检验所用的置换次数设为 200。

我们采用 Gretton 等人（2007）使用的多元基准数据方案。我们针对维度为 p 的随机变量进行独立性检验，其中 $p = 2$、4、10 和 20。数据生成过程如下。首先，使用 R 包 ProDenICA（Hastie 等人 2022）中的 rjordan，从 Gretton 等人（2005b）表3所列的独立成分分析（ICA）基准密度中随机抽取两个单变量随机变量，并进行有放回抽样，生成 n 个观测值；这些基准密度包括超高斯、亚高斯、多峰和单峰分布。其次，使用由角度 θ 参数化的旋转矩阵对这些随机变量进行混合，角度从 0 变化至 $\pi/4$ （角度为零意味着数据相互独立；随着角度增大，依赖关系更容易检测）。也就是说，我们设定 $\theta = 0$、$\pi/8$ 和 $\pi/4$。第三，为每个混合向量添加 $p-1$ 维高斯噪声，其均值为 0，标准差为 1。最后，将每个所得向量分别乘以一个独立的随机 p 维正交矩阵，从而得到在所有观测维度上均相互依赖的向量。我们考虑的样本量分别包括 $n = 30$、50、100、200 和 500。

模拟1中 HSICp、HSICg 和 NEW 检验的经验尺寸和经验功效列于表1。我们可以得出几个有趣的结论。当 $\theta = 0$时，原假设成立，因此我们可以从尺寸控制的角度比较这三种检验的表现。就尺寸控制而言，HSICp 检验和 NEW 检验的水平准确性都非常好，且表现总体相当：在大多数设置下，HSICp 检验和 NEW 检验的经验尺寸通常在 5% 左右，低于 6%，尽管 NEW 检验的经验尺寸略高于 HSICp 检验。诚然，HSICp 检验略优于 NEW 检验，因为

表 1：模拟1中 HSICp、HSICg 和 NEW 检验的经验尺寸和检验功效（以 % 计）。

		$\theta = 0$			$\theta = \pi/8$			$\theta = \pi/4$		
n	p	HSICp	HSICg	NEW	HSICp	HSICg	NEW	HSICp	HSICg	NEW
30	2	5.16	5.31	5.57	8.42	8.74	9.20	13.50	13.69	14.43
	4	5.01	3.49	5.49	6.14	4.12	6.85	7.44	5.19	8.24
	10	5.13	0.10	5.99	5.76	0.06	6.54	5.61	0.14	6.50
	20	5.08	0.00	6.11	4.83	0.00	5.98	5.43	0.00	6.48
50	2	5.08	5.39	5.32	11.84	12.34	12.32	21.66	22.58	22.52
	4	5.22	4.34	5.42	6.67	5.64	7.20	9.59	8.05	10.08
	10	5.33	0.40	5.78	5.52	0.35	5.95	5.47	0.48	5.75
	20	4.86	0.00	5.41	4.77	0.00	5.60	5.44	0.00	6.12
100	2	5.03	5.40	4.97	20.46	21.58	20.74	38.97	40.13	39.38
	4	4.86	4.59	4.91	8.72	8.59	9.24	15.80	15.32	16.27
	10	5.33	1.63	5.51	5.92	1.98	6.29	6.90	2.43	7.15
	20	4.98	0.00	5.28	5.25	0.00	5.69	5.44	0.00	5.72
200	2	5.12	5.54	5.18	38.11	39.78	38.76	58.50	59.83	59.07
	4	5.03	5.27	5.22	15.44	15.99	15.86	30.76	31.25	31.17
	10	5.19	3.10	5.16	6.53	3.93	6.73	9.58	6.27	9.68
	20	4.66	0.03	4.88	5.45	0.00	5.49	6.35	0.12	6.50
500	2	5.19	5.66	5.23	64.05	65.17	64.33	80.60	81.26	80.77
	4	5.03	5.55	5.31	36.04	37.00	36.36	54.99	56.01	55.45
	10	5.25	4.39	5.05	10.62	9.30	10.73	20.11	18.46	20.47
	20	4.64	0.81	4.67	6.15	1.31	6.28	8.78	1.93	8.77

HSICp 检验的经验尺寸介于 4.64% 至 5.33% 之间，而 NEW 检验的经验尺寸介于 4.67% 至 6.11% 之间。然而，HSICg 检验的表现远不如 HSICp 检验和 NEW 检验。当 $p = 2$ 时，HSICg 检验表现相当不错，其经验尺寸通常在 5% 左右。然而，随着维度 p 增大，HSICg 检验变得越来越保守，其经验尺寸会降至 0.00%，尤其是在 $p \geq 10$ 且样本量 n 较小时。这意味着 HSICg 检验不适用于中等或高维度数据，而 HSICp 检验和 NEW 检验仍表现良好。另一方面，当 $\theta > 0$ 时，备择假设成立，因此我们可以从检验功效的角度比较这三种检验的表现。正如预期，HSICp 检验和 NEW 检验的经验功效总体相当，尽管 HSICp 检验的经验功效略低于 NEW 检验；但当维度 $p \geq 10$ 时，HSICg 检验的经验功效通常低于 HSICp 检验和 NEW 检验，这表明了其影响

三种检验的水平准确性。请注意，随着 θ 增大或样本量 n增大，或两者同时增大，三种检验的经验功效通常都会提高。还请注意，随着维度 p 增大，三种检验各自的经验功效都会降低。不过，这并不令人意外，因为当 $\theta \neq 0$时，两个变量中只有第一个元素相关，因此随着维度 p 增大，变量之间的依赖关系会越来越难以检测。

以上，我们从尺寸控制和检验功效方面比较了 HSICp、HSICg 和 NEW 检验。现在我们比较它们的计算成本。为此，表2列出了在 $\theta = 0$、 $p = 4$取10和20，以及 $n = 30$取50、100、200和500时，三种检验进行10, 000次模拟运行的总执行时间（分钟）。可以看出，HSICp 检验的耗时是 NEW 检验的 $10 \sim 100$ 倍，尽管此处置换次数只有200；而对于样本量 $n = 30$取50、100、200和500的情况，NEW 检验的耗时仅约为 HSICg 检验的$1 \sim 10$ 倍。

表2：在模拟1中，$\theta = 0$ 和 $p = 4$，10、和 20 时，HSICp、HSICg 和 NEW 检验在 $N = 10,000$ 次模拟运行中的计算成本（分钟）。

	$p = 4$			$p = 10$			$p = 20$		
n	HSICp	HSICg	NEW	HSICp	HSICg	NEW	HSICp	HSICg	NEW
30	12.75	0.10	0.09	14.61	0.11	0.10	29.41	0.17	0.18
50	40.97	0.18	0.29	46.12	0.20	0.32	86.92	0.36	0.52
100	202.43	0.55	1.51	222.77	0.63	1.61	383.34	1.19	2.41
200	1145.73	1.96	9.32	1227.69	2.29	9.72	1865.75	4.55	12.89
500	13842.41	12.26	123.01	14395.42	14.45	125.72	18071.83	28.38	143.37

3.2 模拟2

在这项模拟研究中，我们从尺寸控制和检验功效两方面比较 HSICp、HSICg 和 NEW 检验在高维数据上的表现。在每次模拟运行中，样本 $(\mathbf{x}_i, \mathbf{y}_i)$ $i = 1, \ldots, n$按如下方式生成。为比较尺寸控制，我们设 $\mathbf{x}_i = \boldsymbol{\Gamma}\mathbf{z}_{1i}, i = 1, \ldots, n$，且 $\mathbf{y}_i = \boldsymbol{\Gamma}\mathbf{z}_{2i}, i = 1, \ldots, n$，其中$\boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top = \boldsymbol{\Sigma}/\mathrm{tr}(\boldsymbol{\Sigma}^2)$，且 $\boldsymbol{\Sigma} = (\sigma_{st})_{s,t=1}^{p}$，$\sigma_{st} = \rho^{|s-t|}$，其中 ρ控制数据相关性。独立同分布随机变量 $\mathbf{z}_{ki} = (z_{ki1}, \ldots, z_{kip})^\top, i = 1, \ldots, n;\ k = 1, 2$ 由以下三个模型生成：

模型 1。 $z_{kij}, j = 1, \ldots, p$ 独立同分布$\sim$ $N(0, 1)$。

模型 2。 $z_{kij} = w_{kij}/\sqrt{2}, \bar{}\ j = 1, \ldots, p$，其中 w_{kij} 独立同分布$\sim$ t_4，服从具有 4 个自由度的 t 分布。

模型 3。 $z_{kij} = (w_{kij} - 1)/\sqrt{2}, \ j = \neg 1, \ldots\ldots, \ p$，其中 w_{kij} 独立同分布$\sim$ χ^2_1。

表3：模拟2中 HSICp、HSICg 和 NEW 检验的经验尺寸（%）。

			$\rho=0.1$			$\rho=0.5$			$\rho=0.9$		
模型	n	p	HSICp	HSICg	NEW	HSICp	HSICg	NEW	HSICp	HSICg	NEW
1	30	50	4.80	0.00	6.10	5.35	0.00	6.57	5.11	1.08	5.56
		100	4.47	0.00	5.85	4.77	0.00	6.02	4.93	0.00	5.53
		200	4.98	0.00	6.14	5.28	0.00	6.35	4.94	0.00	5.65
	50	50	4.88	0.00	5.64	5.24	0.00	5.76	5.14	2.19	5.33
		100	4.74	0.00	5.56	4.59	0.00	5.38	4.91	0.04	5.37
		200	4.83	0.00	5.82	5.06	0.00	5.89	5.19	0.00	5.56
	100	50	5.02	0.00	5.42	4.73	0.00	5.11	5.03	3.64	5.15
		100	4.97	0.00	5.45	4.93	0.00	5.49	4.62	0.58	4.80
		200	5.33	0.00	5.73	5.07	0.00	5.65	4.95	0.00	5.24
2	30	50	5.11	0.00	7.12	5.34	0.00	6.78	5.19	1.44	5.89
		100	4.78	0.00	6.49	5.13	0.00	6.86	5.11	0.02	5.90
		200	4.70	0.00	6.48	5.00	0.00	6.67	4.83	0.00	5.87
	50	50	4.62	0.00	5.80	4.77	0.00	5.79	4.78	2.21	5.30
		100	4.68	0.00	5.82	5.19	0.00	6.09	4.70	0.09	5.27
		200	5.03	0.00	6.02	4.81	0.00	5.75	5.20	0.00	5.81
	100	50	4.83	0.00	5.20	4.81	0.00	5.31	5.25	3.65	5.35
		100	5.03	0.00	5.41	5.27	0.00	5.83	4.96	0.88	5.13
		200	5.19	0.00	5.71	4.71	0.00	5.30	5.01	0.00	5.37
3	30	50	5.04	0.00	6.83	4.42	0.00	5.84	5.10	1.55	5.93
		100	5.17	0.00	6.21	5.08	0.00	6.38	5.31	0.04	6.04
		200	5.31	0.00	6.21	5.11	0.00	6.37	4.76	0.00	5.60
	50	50	4.69	0.00	5.48	5.25	0.00	6.18	4.83	2.40	5.20
		100	5.01	0.00	5.47	4.94	0.00	5.72	4.94	0.17	5.42
		200	5.10	0.00	5.23	4.95	0.00	5.42	4.80	0.00	5.36
	100	50	5.22	0.00	5.39	4.72	0.00	5.07	4.75	3.90	4.86
		100	5.49	0.00	5.71	4.82	0.00	5.05	5.01	1.06	5.17
		200	4.62	0.00	4.45	5.09	0.00	5.34	4.78	0.00	5.11
	ARE		4.04	100.00	16.92	4.07	100.00	17.01	3.09	81.53	9.22

为了进行功效比较，我们令 $\mathbf{x}_i = \boldsymbol{\Gamma}\mathbf{z}_{1i}, i = 1, \ldots, n$，其中 $\mathbf{z}_{1i} = (z_{1i1}, \ldots, z_{1ip})^\top, i = 1, \ldots, n$，且 $z_{1ij}, j = 1, \ldots, p$；$i = 1, \ldots, n$ 由模型 1 生成，并令 $\mathbf{y}_i = (y_{i1}, \ldots, y_{ip})^\top$，其中 $y_{ij} = \delta(x_{ij} + x_{ij}^2) + z_{2ij}$，$j = 1, \ldots, p$；$i = 1, \ldots, n$，且 x_{ij} 为 $\mathbf{x}_i$ 在 $j = 1, \ldots, p$ 中的第 j 个分量；$i = 1, \ldots, n$ 和 $z_{2ij}, j = 1, \ldots, p$；$i = 1, \ldots, n$ 由模型 2 生成。我们设定 $\delta \in \{0.6, 0.8, 1.0, 1.2\}$。注意，随着 δ 增大，$\mathbf{x}_i$ 与 $\mathbf{y}_i$ 之间的依赖关系会增强，因此检验的检验功效应随之提高。我们指定 $\rho = 0.1$、0.5 和 0.9，分别代表低相关、中度相关和高相关数据情形。为了在“大 p、小 n”的设定下生成高维数据，我们分别选择 $n = 30$、50、100 以及 $p = 50$、100、200。此处及下文中，为衡量检验的整体表现，

为保持名义水平 $\alpha = 5\%$，我们采用 Zhang(2011) 提出的平均相对误差（ARE）准则。检验的 ARE 值计算如下：ARE $= 100M^{-1}\sum_{j=1}^{M}|\hat{\alpha}_j - \alpha|/\alpha$，其中 $\hat{\alpha}_j, j = 1, \ldots, M$表示各模拟设置下的经验尺寸。检验的 ARE 值越小，表明该检验在尺寸控制方面的表现越好。

表 3 列出了 HSICp、HSICg 和 NEW 检验在各种设置下的经验尺寸，最后一行给出了它们在 ρ的三个取值下对应的 ARE 值。就尺寸控制而言，无论数据是低相关（$\rho = 0.1$）、中等相关（$\rho = 0.5$）还是高相关 ($\rho = 0.9$)，HSICp 和 NEW 检验的表现都很好，因为它们的经验尺寸通常在 5% 左右，且 ARE 值通常低于 20。它们的表现远优于 HSICg 检验；当 0 时，其经验尺寸为 $\rho = 0.1$ 和 0.5，HSICg 检验完全失效；而当 5% 时，其经验尺寸远低于 $\rho = 0.9$，表现得非常保守。这些结论与模拟1中数据维度较小或中等时观察到的结果一致。

表4展示了不同设置下 HSICp、HSICg 和 NEW 检验的经验功效。可以看出，在固定 n 和 p的情况下，随着 δ 的值增大，三种检验的经验功效均会提高；在每种设置下，NEW 检验的经验功效通常大于 HSICp 检验，而 HSICg 检验则完全没有检验功效。从表3可以明显看出，三种检验的功效大小顺序会受到其经验尺寸的影响。

3.3 模拟3

在本模拟研究中，我们将 NEW 检验与几种有代表性的现有检验进行比较，以解决函数型数据的两样本独立性检验问题。这些现有检验包括基于皮尔逊相关的检验、基于动态相关的检验（Dubin 和 Müller（2005））、基于FPCA距离协方差的检验（Kosorok（2009））以及基于全局时间相关的检验（Zhou 等人（2018））, 分别记为 Pearson、dnm、FPCA 和 gtemp。dnm、FPCA 和 gtemp 检验均基于置换，并在Miao等 (2022)中实现。

在每次模拟运行中，函数型观测 $\{x_i(t)\ y_i(t)\}, t \in [0, 1], i = 1, \ldots, n$ 按如下方式生成。令 $x_i(t) = \sum_{j=1}^{50} z_{1ij}\sqrt{2}\cos(j\pi t)$，且 $y_i(t) = \sum_{j=1}^{50} z_{2ij}\sqrt{2}\cos(j\pi t)$，其中 $z_{1ij}, j = 1, \ldots, 50$； $i = 1, \ldots, n$ 是按模拟2中所述相同方式生成的独立同分布随机变量。对于某个给定函数 $f(\cdot)$，令 $z_{2ij} = f(z_{1ij})$，其中 $j = 1, \ldots, m$，并令 z_{2ij} 为独立同分布的 $\sim N(0, 1)$，其中 $j = m+1, \ldots, 50$。这里， m 是一个用于控制函数型观测依赖程度的整数：当 $m = 0$ 时，原假设成立；否则，依赖程度随 m 增大。在数值实现中，将在一组等间距时间点上对这些函数型观测求值： $t_r = (r-1)/(k-1)\ r = 1, \ldots, k$。为考察检验功效，我们考虑四个函数：$f(u) = u^3$、$f(u) = u^2$、$f(u) = u\sin(u)$，以及 $f(u) = u\cos(u)$；第一个函数是单调的，其余三个则不是。详情见图1。当

表4：模拟2中 HSICp、HSICg 和 NEW 检验的经验功效（%）。

			$\rho = 0.1$			$\rho = 0.5$			$\rho = 0.9$		
n	p	δ	HSICp	HSICg	NEW	HSICp	HSICg	NEW	HSICp	HSICg	NEW
30	50	0.6	15.29	0.00	18.22	15.54	0.00	18.21	16.11	0.00	18.51
		0.8	27.17	0.00	31.47	29.05	0.00	33.05	29.42	0.01	32.63
		1.0	45.64	0.00	50.97	47.45	0.00	52.65	49.11	0.01	52.60
		1.2	66.95	0.00	71.14	68.81	0.00	73.15	70.86	0.09	73.93
	100	0.6	10.42	0.00	13.02	10.27	0.00	12.96	11.21	0.00	13.49
		0.8	16.17	0.00	19.39	16.42	0.00	20.18	17.48	0.00	20.50
		1.0	26.27	0.00	30.63	26.53	0.00	31.17	28.84	0.00	32.58
		1.2	39.93	0.00	45.16	42.59	0.00	47.70	45.22	0.00	49.55
	200	0.6	8.06	0.00	9.89	8.43	0.00	10.31	7.76	0.00	9.74
		0.8	10.58	0.00	13.03	11.03	0.00	13.79	10.97	0.00	13.65
		1.0	15.75	0.00	19.08	15.49	0.00	18.66	17.22	0.00	20.18
		1.2	23.53	0.00	27.44	22.99	0.00	27.53	25.06	0.00	29.26
50	50	0.6	26.75	0.00	29.53	26.48	0.00	28.82	27.93	0.04	29.77
		0.8	52.32	0.00	55.93	54.71	0.00	57.84	55.54	0.09	57.62
		1.0	81.09	0.00	83.12	82.04	0.00	84.24	81.72	0.30	83.23
		1.2	95.68	0.00	96.41	96.74	0.00	97.18	95.96	0.97	96.36
	100	0.6	15.48	0.00	17.57	15.36	0.00	17.29	16.61	0.00	18.47
		0.8	28.26	0.00	31.42	29.38	0.00	32.48	32.07	0.00	34.43
		1.0	49.88	0.00	53.68	51.26	0.00	55.24	54.97	0.00	57.73
		1.2	72.03	0.00	74.95	76.19	0.00	78.89	78.81	0.01	80.73
	200	0.6	9.84	0.00	11.49	10.38	0.00	12.05	10.23	0.00	11.74
		0.8	16.20	0.00	18.30	16.66	0.00	18.98	17.56	0.00	19.90
		1.0	26.55	0.00	29.66	27.42	0.00	30.54	30.16	0.00	33.10
		1.2	42.78	0.00	46.14	44.97	0.00	48.83	47.36	0.00	50.70
100	50	0.6	62.82	0.00	64.73	64.45	0.00	66.01	63.28	0.25	64.19
		0.8	95.04	0.00	95.65	95.64	0.00	95.95	94.30	1.05	94.55
		1.0	99.93	0.00	99.95	99.99	0.07	99.99	99.78	3.19	99.81
		1.2	100.00	0.00	100.00	100.00	0.15	100.00	100.00	15.96	100.00
	100	0.6	33.23	0.00	35.29	34.88	0.00	36.72	36.05	0.00	37.60
		0.8	67.56	0.00	69.55	69.28	0.00	71.21	72.47	0.03	73.59
		1.0	93.29	0.00	93.94	94.91	0.00	95.55	95.23	0.14	95.59
		1.2	99.64	0.00	99.68	99.75	0.00	99.78	99.81	0.22	99.85
	200	0.6	18.29	0.00	19.64	18.07	0.00	19.33	20.01	0.00	21.32
		0.8	35.98	0.00	38.15	37.69	0.00	40.10	39.87	0.00	41.77
		1.0	63.61	0.00	65.94	65.66	0.00	67.87	69.62	0.00	71.48
		1.2	87.67	0.00	89.02	89.35	0.00	90.67	92.13	0.01	92.92

当 $f(u) = u\sin(u)$ 和 $f(u) = u\cos(u)$ 时，数值大于 $f(u) = u^3$ 和 $f(u) = u^2$ 时的情形。因此，当 $f(u) = u^3$ 和 $f(u) = u^2$ 时，我们设定 $m = 3$，5，10 和 15；当 $f(u) = u\sin(u)$ 和 $f(u) = u\cos(u)$ 时，则设定 $m = 15$，20，25 和 30。

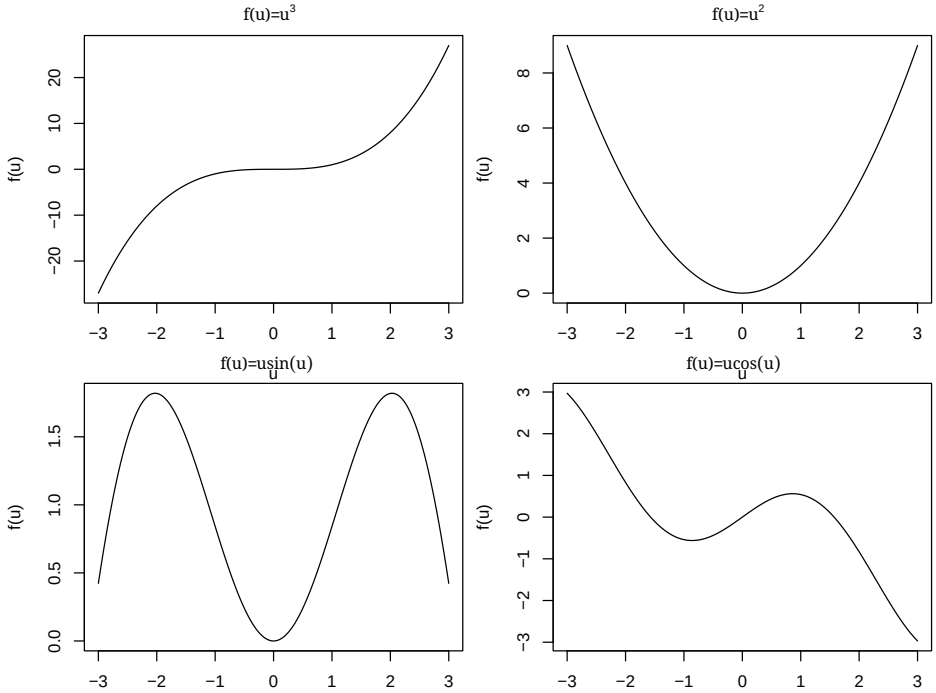


图1： 模拟*3*中考虑的四个函数。

表 5 展示了所考虑的五种检验的经验尺寸，最后一行是相应的渐近相对效率值（ARE）。可以看出，这五种检验都具有良好的水平准确性，其经验尺寸通常在5%左右。诚然，就尺寸控制而言，NEW 检验的表现略逊于其他四种检验，但其计算速度也明显快于它们。

为节省篇幅，表 6 仅展示了模型 1 在 $k = 201$ 下五种所考虑检验的经验功效（%），因为针对 k 其他取值所得出的结论类似。这些经验功效在几个方面都颇具揭示性。首先，对于单调关系（即当 $f(u) = u^3$ 时），NEW 检验的检验功效仅略低于其他四种检验。NEW 检验检验功效略低的原因可能在于，NEW 检验是基于 HSIC 的检验；正如 Shen 等人（2019）所指出的，对于检测单调关系，基于希尔伯特-施密特独立性准则的检验的表现可能略逊于基于距离协方差的检验。其次，对于非单调关系（即当 $f(u) = u^2$、$u\sin(u)$ 以及 $u\cos(u)$ 时），NEW 检验通常比其他四种检验具有更高的检验功效。对于不同的非单调关系，其他四种检验的表现差异很大。例如，当 $f(u) = u^2$ 且 $u\sin(u)$ 时，Pearson 检验几乎没有检验功效；当 $f(u) = u\sin(u)$ 且 $u\cos(u)$ 时，gtemp 检验几乎没有检验功效；而当 $f(u) = u\sin(u)$ 时，dnm 和函数主成分分析检验的检验功效都非常低。

表5：模拟3中皮尔逊、dnm、函数主成分分析、gtemp 和 NEW 检验的经验尺寸（%）。

模型	n	k	皮尔逊	dnm	FPCA	gtemp	NEW
1	30	201	4.89	5.12	4.68	5.02	6.88
		501	5.13	5.08	4.42	5.19	6.09
		1001	5.17	5.39	4.70	4.96	6.13
	50	201	4.83	4.94	4.36	5.21	5.69
		501	4.92	4.86	4.56	5.08	5.98
		1001	5.02	4.90	4.55	4.89	5.92
	100	201	4.84	5.07	4.34	5.01	5.45
		501	4.84	4.91	4.59	5.14	5.72
		1001	4.79	5.04	4.78	5.13	5.48
2	30	201	4.55	4.73	4.79	4.77	6.29
		501	5.06	4.65	4.26	5.19	6.50
		1001	5.11	4.89	4.77	4.85	6.21
	50	201	4.79	4.86	4.92	4.88	6.25
		501	4.93	4.73	4.35	4.83	5.62
		1001	5.52	5.35	4.37	4.95	6.31
	100	201	5.26	5.33	4.41	5.37	5.08
		501	5.13	5.00	4.65	5.13	5.31
		1001	4.84	4.99	4.47	4.57	5.84
3	30	201	5.27	5.17	4.62	4.79	6.17
		501	5.36	5.29	4.09	4.93	6.23
		1001	5.07	4.98	4.40	5.16	6.28
	50	201	5.11	5.04	4.57	5.03	5.95
		501	5.13	5.03	4.96	5.25	6.06
		1001	4.78	4.83	4.62	4.96	5.39
	100	201	4.96	5.20	4.52	4.94	5.33
		501	5.55	5.65	4.46	5.15	5.51
		1001	4.87	4.91	4.34	5.01	5.41
	ARE		3.75	3.39	9.22	2.78	17.84

综合上述三项模拟研究，就水平准确性、检验功效和计算成本而言，NEW 检验总体上优于其他竞争方法，因此应推荐将其用于真实数据分析。

4 函数数据和高维数据的应用

本节中，我们将介绍 NEW 检验及上一节提到的几种现有竞争方法在一个函数型数据集和一个高维数据集上的应用。对于

表 6：模型 1 下模拟3中皮尔逊、dnm、函数主成分分析、gtemp 和 NEW 检验的经验功效（%），其中 $k = 201$。

		$f(u) = u^3$					$f(u) = u^2$				
n	m	皮尔逊	dnm	FPCA	gtemp	NEW	皮尔逊	dnm	FPCA	gtemp	NEW
30	3	91.32	95.98	83.51	75.89	54.25	4.90	8.25	8.03	5.98	9.68
	5	99.90	99.97	97.07	99.04	79.72	5.28	9.98	10.96	6.47	12.83
	10	100	100	99.98	100	97.62	5.01	14.19	18.25	8.45	23.33
	15	100	100	99.99	100	99.65	4.49	17.41	26.60	9.22	37.48
50	3	99.19	99.73	99.28	96.00	79.70	4.80	7.59	8.90	5.83	11.02
	5	100	100	100	99.99	97.47	5.18	10.17	11.89	7.03	15.70
	10	100	100	100	100	99.99	4.89	14.61	20.87	8.34	31.80
	15	100	100	100	100	100	4.54	16.96	32.32	10.16	53.21
100	3	100	100	100	99.99	99.16	4.83	7.77	9.54	5.63	14.37
	5	100	100	100	100	100	4.73	9.38	14.02	6.60	23.82
	10	100	100	100	100	100	4.84	13.62	27.72	8.38	57.57
	15	100	100	100	100	100	4.73	17.70	42.68	10.40	85.17
		$f(u) = u$正弦(u)					$f(u) = u$余弦(u)				
n	m	皮尔逊	dnm	FPCA	gtemp	NEW	皮尔逊	dnm	FPCA	gtemp	NEW
30	15	4.59	6.42	5.53	4.91	9.63	5.44	8.53	6.90	5.14	9.46
	20	5.24	6.95	6.14	5.64	11.35	6.22	11.12	7.84	6.12	11.15
	25	4.82	7.77	6.95	5.93	14.07	6.67	14.16	9.85	6.53	14.42
	30	4.71	9.42	8.27	6.13	17.02	8.81	19.15	11.86	7.51	18.55
50	15	5.13	6.82	5.84	5.35	10.08	5.71	9.28	6.73	5.74	9.74
	20	5.00	7.65	6.52	5.67	13.44	7.89	13.35	8.08	6.26	12.75
	25	5.07	8.55	7.79	5.36	17.92	9.12	17.51	10.04	6.66	16.95
	30	5.09	9.44	8.31	6.29	24.27	11.49	23.39	11.12	7.20	23.15
100	15	5.03	6.62	6.47	5.24	15.33	7.19	10.89	6.82	5.49	12.11
	20	5.38	7.46	7.10	4.96	21.69	9.60	16.10	7.77	6.13	16.71
	25	4.84	8.93	8.38	5.83	32.36	13.27	23.21	9.55	6.63	25.10
	30	4.94	9.56	9.51	5.76	44.74	18.38	33.68	11.40	7.31	36.51

NEW 检验，我们继续使用高斯径向基函数核，并按照上一节所述选择核宽度。

4.1 加拿大气象数据

在本小节中，我们使用加拿大天气数据集，将皮尔逊、dnm、函数主成分分析、gtemp 和 NEW 检验应用于函数型数据，数据集已在第 1 节中作了简要介绍。对于 35 个气象站中的每一个，在为期一年的时间内，变量“温度”记录日平均温度，变量“降水”记录四舍五入到 0.1 毫米的日平均降雨量。原始温度

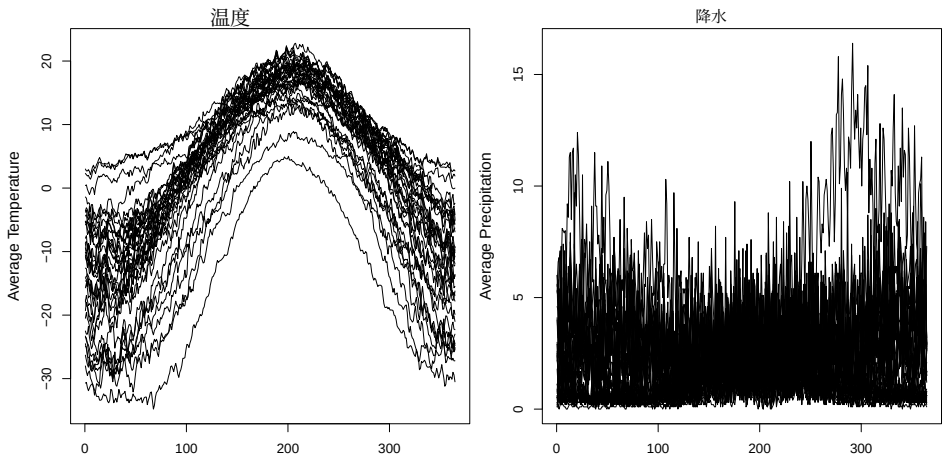


图 2： 加拿大35 个气象站的原始温度曲线和降水曲线。

和降水曲线见图 35 个气象站的曲线如图 2 所示。由于日平均温度和日平均降水量来自同一批 35 个加拿大气象站，预计两者之间存在一定的依赖关系。我们关注的是这种依赖关系有多强。

为此，我们将皮尔逊、dnm、函数主成分分析、gtemp 和 NEW 检验应用于这组加拿大天气数据集，以检验温度曲线与降水曲线是否相互独立。五种检验的 p值如表 7所示。可以看出，五种检验的 p值都相当小，远低于 1%，这表明有充分证据拒绝原假设；也就是说，正如预期的那样，35个加拿大气象站的温度曲线与降水曲线之间存在很强的依赖关系。

表 7：用于检验潜在温度曲线与潜在降水曲线之间独立性的皮尔逊、dnm、函数主成分分析、gtemp 和 NEW 检验的 p 值。

皮尔逊	dnm	FPCA	gtemp	NEW
2.73×10^{-3}	0	5.00×10^{-3}	0	8.29×10^{-7}

4.2 结肠数据

在本小节中，我们使用著名的结肠数据集，演示希尔伯特-施密特独立性准则p、HSICg 和 NEW 检验在高维数据中的应用。该数据集包含 62 个组织，每个组织包含 2000 个基因表达水平，可从以下网址下载：http://microarray.princeton.edu/oncology/affydata/index.html。为了构造高维数据的双样本独立性检验，我们选取前 31 个组织组成第一组，其余 31 个组织组成第二组。因此，由于这些组织相互独立，两组也应当独立。表 8 列出了所考虑的三种检验的 p值

这些值均大于 50%，表明完全没有证据可以拒绝原假设，这与预期一致。

表8：HSICp、HSICg 和 NEW 检验用于检验结肠数据两组之间独立性的 p 值。

HSICp	HSICg	NEW
0.956	0.619	0.560

为进一步展示 NEW 检验相较于 HSICp 和 HSICg 检验的水平准确性，我们基于该结肠数据集进行了小规模模拟研究，以模拟三种检验的经验尺寸；模拟结果来自 10,000 次模拟运行。每次运行中，将 62 个组织随机分成两个大小相等的组。根据 p 值小于名义水平 $\alpha = 5\%$ 或 10% 的次数占 10,000 次运行的比例，计算检验的经验尺寸。三种检验的经验尺寸列于表 9。可以看出，NEW 和 HSICp 检验的水平准确性都很好，而 HSICg 检验则相当保守。这与表 1 和 3中模拟结果所得到的结论一致。

表9：通过小规模模拟研究获得的 HSICp、HSICg 和 NEW 检验的经验尺寸（%）。

α	HSICp	HSICg	NEW
5%	4.96	2.25	6.80
10%	9.76	5.71	11.60

5 总结

文献中已有多种基于希尔伯特–施密特独立性准则（HSIC）的可分度量空间两样本独立性检验。本文提出并研究了一种新的、基于 HSIC 的可分度量空间独立性检验，并将其应用于函数型数据和高维数据。在某些正则条件和原假设下，可以证明所提出的检验统计量渐近服从卡方型混合极限。为实施该检验，我们采用 Zhang (2005) 的三累积量匹配卡方近似方法，并使用数据一致地估计近似参数，以近似卡方混合分布。模拟研究和真实数据应用表明，就尺寸控制、检验功效和计算成本而言，所提出的检验在多变量、高维和函数型数据方面优于若干现有检验。然而，表 3和 5 也表明，当样本量较小时，所提出的检验仍略显宽松。进一步提高该检验水平准确性的方法值得研究，也有其必要性。

附录

~ ~引理 1 的证明。根据 $K(x_i, x_j)$ 在式（6）中的表达式，以及 $K^*(x_i, x_j)$ 在式（7）中的表达式，对于任意固定的 $i, j \in \{1, \ldots, n\}$，有

$$|\tilde{K}^*(x_i, x_j) - \tilde{K}(x_i, x_j)| \leq |n^{-1}(n-1)^{-1} \sum_{u \neq v} K(x_u, x_v) - \mathrm{E}_{z,z'}[K(z, z')]|$$
$$+ |(n-1)^{-1} \sum_{v \neq i} K(x_i, x_v) - \mathrm{E}_{z'}[K(x_i, z')]| + |(n-1)^{-1} \sum_{u \neq j} K(x_u, x_j) - \mathrm{E}_z[K(z, x_j)]|.$$

由此可得，当 $n \to \infty$ 时，我们有

$$\mathrm{E}\{n[n^{-1}(n-1)^{-1} \sum_{u \neq v} K(x_u, x_v) - \mathrm{E}_{z,z'}[K(z, z')]]\} = o(1), \quad \text{and}$$
$$\mathrm{Var}\{n[n^{-1}(n-1)^{-1} \sum_{u \neq v} K(x_u, x_v) - \mathrm{E}_{z,z'}[K(z, z')]]\} \leq \mathrm{E}[K(x, x')]^2 \leq 16 B_K^2.$$

因此，当 $n \to \infty$ 时，我们有 $|n^{-1}(n-1)^{-1} \sum_{u \neq v} K(x_u, x_v) - \mathrm{E}_{z,z'}[K(z, z')]| = \mathcal{O}(n^{-1})$。类似地，我们有 $|(n-1)^{-1} \sum_{v \neq i} K(x_i, x_v) - \mathrm{E}_{z'}[K(x_i, z')]| = \mathcal{O}(n^{-1/2})$，以及 $|(n-1)^{-1} \sum_{u \neq j} K(x_u, x_j) - \mathrm{E}_z[K(z, x_j)]| = \mathcal{O}(n^{-1/2})$。因此，当 $n \to \infty$ 时，我们有

$$\tilde{K}^*(x_i, x_j) = \tilde{K}(x_i, x_j) + \mathcal{O}(n^{-1/2}) \text{ uniformly for all } x_i, x_j.$$

引理得证。 □

定理 1的证明。在原假设下，x 和 y 相互独立。然后，在条件（11）下，根据（12），我们有

$$\mathrm{E}[\tilde{K}(x, x)\tilde{L}(y, y)] = \mathrm{E}[\tilde{K}(x, x)]\,\mathrm{E}[\tilde{L}(y, y)] = \left(\sum_{r=1}^{\infty} \lambda_r\right)\left(\sum_{s=1}^{\infty} \rho_s\right) < \infty,$$

and

$$\mathrm{E}[\tilde{K}(x, x')\tilde{L}(y, y')]^2 = \mathrm{E}[\tilde{K}(x, x')]^2\,\mathrm{E}[\tilde{L}(y, y')]^2 = \left(\sum_{r=1}^{\infty} \lambda_r^2\right)\left(\sum_{s=1}^{\infty} \rho_s^2\right) < \infty.$$

此外，我们有

$$\begin{aligned} \tilde{K}(x, x')\tilde{L}(y, y') &= \sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \lambda_r \rho_s \phi_r(x)\phi_r(x')\psi_s(y)\psi_s(y') \\ &= \sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \lambda_r \rho_s [\phi_r(x)\psi_s(y)][\phi_s(x')\psi_s(y')], \end{aligned}$$

其中 (x', y') 是 (x, y) 的独立副本。由于 x 和 y 独立，我们可以证明 $\phi_r(x)\psi_s(y)$ 是标准正交的。事实上，我们有

$$\mathrm{E}\{[\phi_r(x)\psi_s(y)][\phi_\alpha(x)\psi_\beta(y)]\} = \mathrm{E}\{[\phi_r(x)\phi_\alpha(x)][\psi_s(y)\psi_\beta(y)]\}$$
$$= \mathrm{E}[\phi_r(x)\phi_\alpha(x)]\,\mathrm{E}[\psi_s(y)\psi_\beta(y)] = \delta_{r\alpha}\delta_{s\beta},$$

其取值为 1，当 $(r, s) = (\alpha, \beta)$；否则取值为 0。由此可知，$\lambda_r \rho_s$，$r, s = 1, 2, \ldots\ldots$是 $K(x, x')L(y, y')$ 的特征元~ ~，对应的特征函数为 $\phi_r(x)\psi_s(y)$ $r, s = 1, 2, \ldots\ldots$。

~我们可以将 T_n 改写为

$$\tilde{T}_n = n^{-1}\sum_{i=1}^{n}[\tilde{K}(x_i,x_i)\tilde{L}(y_i,y_i)] + (n-1)\binom{n}{2}^{-1}\sum_{1\le i<j\le n}[\tilde{K}(x_i,x_j)\tilde{L}(y_i,y_j)].$$

根据大数定律，当 $n\to\infty$ 时，我们有 $n^{-1}\sum^{n}[\tilde{K}(x_i,x_i)\tilde{L}(y_i,y_i)] \xrightarrow{a.s.} \mathrm{E}[\tilde{K}(x,x)\tilde{L}(y,y)] = \left(\sum_{r=1}^{\infty}\lambda_r\right)\left(\sum_{s=1}^{\infty}\rho_s\right)$。此外，由于

$$\mathrm{E}_{x,y}[\tilde{K}(x,x')\tilde{L}(y,y')] = \mathrm{E}_{x,y'}[\tilde{K}(x,x')\tilde{L}(y,y')] = \mathrm{E}_{x,y,x',y'}[\tilde{K}(x,x')\tilde{L}(y,y')] = 0$$

且 $\mathrm{E}[\tilde{K}(x,x')\tilde{L}(y,y')]^2 = \left(\sum_{r=1}^{\infty}\lambda_r^2\right)\left(\sum_{s=1}^{\infty}\rho_s^2\right) < \infty$，根据 Serfling（1980，第 194 页）的 U 统计量定理，当 $n\to\infty$ 时，我们有 $(n-1)\binom{n}{2}^{-1}\sum_{1\le i<j\le n}[\tilde{K}(x_i,x_j)\tilde{L}(y_i,y_j)] \xrightarrow{\mathcal{L}} \sum_{r=1}^{\infty}\sum_{s=1}^{\infty}\lambda_r\rho_s(A_{rs}-1)$ A_{rs} 独立同分布$\sim$ χ_1^2.

由此可得 $\tilde{T}_n \xrightarrow{\mathcal{L}} \tilde{T}$ 其中 $\tilde{T} = \sum_{r=1}^{\infty}\sum_{s=1}^{\infty}\lambda_r\rho_s + \sum_{r=1}^{\infty}\sum_{s=1}^{\infty}\lambda_r\rho_s(A_{rs}-1)$

$= \sum_{r=1}^{\infty}\sum_{s=1}^{\infty}\lambda_r\rho_s A_{rs}$, A_{rs} 独立同分布$\sim$ χ_1^2

根据引理1，我们有 $T_n = \tilde{T}_n + \mathcal{O}(n^{-1/2})$。这意味着 T_n 和 $\tilde{T}_n$ 具有相同的渐近分布。也就是说，$T_n \xrightarrow{\mathcal{L}} \tilde{T}$ 其中 $\tilde{T}$ 如上所定义。 □

定理 2 的证明。根据引理 1，我们有 $T_n = \tilde{T}_n + \mathcal{O}(n^{-1/2})$。由此可得 $\mathrm{E}(T_n) = \mathrm{E}(\tilde{T}_n) + \mathcal{O}(n^{-1/2})$、$\mathrm{Var}(T_n) = \mathrm{Var}(\tilde{T}_n) + \mathcal{O}(n^{-1/2})$，以及 $\mathrm{E}[T_n - \mathrm{E}(T_n)]^3 = \mathrm{E}[\tilde{T}_n - \mathrm{E}(\tilde{T}_n)]^3 + \mathcal{O}(n^{-1/2})$。因此，我们只需求出 $\mathrm{E}(\tilde{T}_n)$、$\mathrm{Var}(\tilde{T}_n)$ 和 $\mathrm{E}[\tilde{T}_n - \mathrm{E}(\tilde{T}_n)]^3$。注意

$$\tilde{T}_n = n^{-1}\sum_{i=1}^{n}[\tilde{K}(x_i,x_i)\tilde{L}(y_i,y_i)] + 2n^{-1}\sum_{1\le i<j\le n}[\tilde{K}(x_i,x_j)\tilde{L}(y_i,y_j)].$$

~由于 $\mathrm{E}_x[K(x,x')] = \mathrm{E}_{x'}[K(x,x')] = \mathrm{E}_{x,x'}[K(x,x')] = 0$ 且 $\mathrm{E}_y[L(y,y')] = \mathrm{E}_{y'}[L(y,y')] = \mathrm{E}_{y,y'}[K(y,y')] = 0$，因此我们有

$$\mathrm{E}(\tilde{T}_n) = n^{-1}\sum_{i=1}^{n}\mathrm{E}[\tilde{K}(x_i,x_i)\tilde{L}(y_i,y_i)] = \mathrm{E}[\tilde{K}(x,x)]\,\mathrm{E}[\tilde{L}(y,y)] = M_1N_1.$$

由此可得

$$\begin{aligned}\tilde{T}_n - \mathrm{E}(\tilde{T}_n) &= n^{-1}\sum_{i=1}^{n}\{\tilde{K}(x_i,x_i)\tilde{L}(y_i,y_i) - M_1N_1\} + 2n^{-1}\sum_{1\le i<j\le n}[\tilde{K}(x_i,x_j)\tilde{L}(y_i,y_j)]\\ &= n^{-1}\sum_{i=1}^{n}A_i + 2n^{-1}\sum_{1\le i<j\le n}B_{ij}.\end{aligned}$$

Not使得 $\mathrm{E}(A_i) = 0$ 和 A_1，A_2，$\cdots$，A_n 为独立同分布变量，并且在原假设下，我们有

$$\mathrm{E}_{x_iy_i}(B_{ij}) = \mathrm{E}_{x_jy_j}(B_{ij}) = \mathrm{E}(B_{ij}) = 0.$$

Thus

$$\begin{aligned}&\mathrm{Var}(\tilde{T}_n) = n^{-1}\,\mathrm{Var}(A_1) + 4n^{-2}\sum_{1\le i<j\le n}\mathrm{E}(B_{ij}^2)\\ =\;& n^{-1}\,\mathrm{Var}[\tilde{K}(x,x')]\,\mathrm{Var}[\tilde{L}(y,y')] + 2(1-n^{-1})\,\mathrm{E}[\tilde{K}^2(x,x')]\,\mathrm{E}[\tilde{L}^2(y,y')]\\ =\;& 2M_2N_2 + \mathcal{O}(n^{-1}).\end{aligned}$$

最后，

$$\begin{aligned}\mathrm{E}[\tilde{T}_n - \mathrm{E}(\tilde{T}_n)]^3 &= n^{-3}\,\mathrm{E}\left(\sum_{i=1}^{n} A_i\right)^3 + 8n^{-3}\,\mathrm{E}\left(\sum_{1\le i<j\le n} B_{ij}\right)^3 \\ &= n^{-2}\,\mathrm{E}(A_1^3) + 8n^{-3}\,\mathrm{E}\left(\sum_{1\le i<j\le n} B_{ij}\right)^3,\end{aligned}$$

其中 $\mathrm{E}(A_1^3) = \mathrm{E}\left[\tilde{K}(x,\ x)\tilde{L}(y,\ y) - M_1N_1\right]^3$ 且

$$\begin{aligned}\mathrm{E}\left(\sum_{1\le i<j\le n} B_{ij}\right)^3 &= \mathrm{E}\left[\sum_{i<j} B_{ij}^3 + 3\sum\nolimits^{*} B_{ij}^2 B_{\alpha\beta} + 6\sum\nolimits^{**} B_{ij}B_{\alpha\beta}B_{uv}\right] \\ &= \left\{\tfrac{n(n-1)}{2}\,\mathrm{E}[B_{12}^3] + 6\tfrac{n(n-1)(n-2)}{3!}\,\mathrm{E}[B_{12}B_{23}B_{31}]\right\} \\ &= n(n-1)(n-2)M_3N_3 + \tfrac{n(n-1)}{2}\,\mathrm{E}[\tilde{K}^3(x,x')]\,\mathrm{E}[\tilde{L}^3(y,y')],\end{aligned}$$

其中 $*$ 表示“$i<j,\ \alpha<\beta$”和“$(i,\ j)\neq(\alpha,\ \beta)$”，而 $**$ 表示“$i<j,\ \alpha<\beta,\ u<v$”和“$(i,\ j),\ (\alpha,\ \beta),\ (r,\ s)$ 彼此不相等。”因此 $\mathrm{E}[\tilde{T}_n - \mathrm{E}(\tilde{T}_n)]^3 = 8M_3N_3 + \mathcal{O}(n^{-1})$。定理得证。

□

证明（*19*）。注意到

$$\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}[\tilde{K}^*(x_i,x_j)]^2 - \sum_{i=1}^{n}[\tilde{K}^*(x_i,x_i)]^2\right\} = \mathrm{tr}(\tilde{\boldsymbol{K}}^{*2}) - \mathrm{tr}(\tilde{\boldsymbol{K}}^{*}o\tilde{\boldsymbol{K}}^{*}).$$

由此可得 $\hat{}\ M_2 = [n(n-1)]^{-1}[\mathrm{tr}(\tilde{}\boldsymbol{K}^{*2}) - \mathrm{tr}(\tilde{}\boldsymbol{K}^{*}o\tilde{}\boldsymbol{K}^{*})]$。此外，我们有

$$\begin{aligned}&\sum_{1\le i<j<k\le n}[\tilde{K}^*(x_i,x_j)\tilde{K}^*(x_j,x_k)\tilde{K}^*(x_k,x_i)] \\ =\ &\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\ell=1}^{n}\tilde{K}^*(x_i,x_j)\tilde{K}^*(x_j,x_k)\tilde{K}^*(x_k,x_i) - \sum_{i=1}^{n}[\tilde{K}^*(x_i,x_i)]^3 - 3\sum_{i\neq j}\tilde{K}^*(x_i,x_j)\tilde{K}^*(x_j,x_i)\tilde{K}^*(x_i,x_i) \\ =\ &\mathrm{tr}(\tilde{\boldsymbol{K}}^{*3}) - \mathrm{tr}(\tilde{\boldsymbol{K}}^{*}o\tilde{\boldsymbol{K}}^{*}o\tilde{\boldsymbol{K}}^{*}) - 3\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{K}^*(x_i,x_j)\tilde{K}^*(x_j,x_i)\tilde{K}^*(x_i,x_i) - \sum_{i=1}^{n}[\tilde{K}^*(x_i,x_j)]^3\right\} \\ =\ &\mathrm{tr}(\tilde{\boldsymbol{K}}^{*3}) - \mathrm{tr}(\tilde{\boldsymbol{K}}^{*}o\tilde{\boldsymbol{K}}^{*}o\tilde{\boldsymbol{K}}^{*}) - 3\left\{\mathrm{tr}[\mathrm{diag}(\tilde{\boldsymbol{K}}^{*})\tilde{\boldsymbol{K}}^{*2}] - \mathrm{tr}(\tilde{\boldsymbol{K}}^{*}o\tilde{\boldsymbol{K}}^{*}o\tilde{\boldsymbol{K}}^{*})\right\} \\ =\ &\mathrm{tr}(\tilde{\boldsymbol{K}}^{*3}) - 3\,\mathrm{tr}[\mathrm{diag}(\tilde{\boldsymbol{K}}^{*})\tilde{\boldsymbol{K}}^{*2}] + 2\,\mathrm{tr}(\tilde{\boldsymbol{K}}^{*}o\tilde{\boldsymbol{K}}^{*}o\tilde{\boldsymbol{K}}^{*}).\end{aligned}$$

由此可得 $\hat{}\ M_3 = [n(n-1)(n-2)]^{-1}\left\{\mathrm{tr}(\tilde{}\boldsymbol{K}^{*3}) - 3\,\mathrm{tr}[\mathrm{diag}(\tilde{}\boldsymbol{K}^{*})\ \tilde{}\boldsymbol{K}^{*2}] + 2\,\mathrm{tr}(\tilde{}\boldsymbol{K}^{*}o\tilde{}\boldsymbol{K}^{*}o\tilde{}\boldsymbol{K}^{*})\right\}$。

□

定理 3 的证明。在条件 (11) 下，由引理 1 可得，对于所有 x_i, x_j，$\tilde{K}^*(x_i,x_j) = \tilde{K}(x_i,x_j) + \mathcal{O}(n^{-1/2})$ 一致成立。由于对于所有 $x, x' \in \mathcal{X}$，$|K(x,x')| \le 4B_K < \infty$，根据 (18)，我们有

$$\hat{M}_1 = \tilde{M}_1 + \mathcal{O}(n^{-1/2}),\ \ \hat{M}_2 = \tilde{M}_2 + \mathcal{O}(n^{-1/2}),\quad \text{and}\quad \hat{M}_3 = \tilde{M}_3 + \mathcal{O}(n^{-1/2}),$$

其中

$$\begin{aligned}\tilde{M}_1 = \tfrac{1}{n}\sum_{i=1}^{n}\tilde{K}(x_i,x_i),\ \ \tilde{M}_2 = \tfrac{2}{n(n-1)}\textstyle\sum_{1\le i<j\le n}\tilde{K}^2(x_i,x_j), \\ \tilde{M}_3 = \tfrac{6}{n(n-1)(n-2)}\textstyle\sum_{1\le i<j<k\le n}\tilde{K}(x_i,x_j)\tilde{K}(x_j,x_k)\tilde{K}(x_k,x_i).\end{aligned}$$

˜ ˜ ˜由于 M_1， M_2， M_3 分别是关于 M_1， M_2， M_3 的 U 统计量，且在条件（11）下，我们有

$$\mathrm{E}[\tilde{K}(x,x)]^2 \leq (4B_K)^2 < \infty, \ \mathrm{E}\left[\tilde{K}(x,x')\right]^4 \leq (4B_K)^4 < \infty,$$
$$\mathrm{E}\left[\tilde{K}(x,x')\tilde{K}(x',x'')\tilde{K}(x'',x)\right]^2 \leq (4B_K)^6 < \infty.$$

由此，根据（Serfling，1980，第185页）的引理 A，当 $n \to \infty$ 时，我们有 $\tilde{M}_1 \xrightarrow{p} M_1$、$\tilde{M}_2 \xrightarrow{p} M_2$ 和 $\tilde{M}_3 \xrightarrow{p} M_3$。因此，当 $n \to \infty$ 时，我们有 $\hat{}M_1 \xrightarrow{p} M_1$、$\tilde{M}_2 \xrightarrow{p} M_2$ 和 $\tilde{M}_3 \xrightarrow{p} M_3$。于是，当 $n \to \infty$ 时，我们有 $\hat{}M_\ell \xrightarrow{p} M_\ell$、$\ell = 1$、2、3。类似地，可以证明，当 $n \to \infty$ 时，我们有 $\hat{}N_\ell \xrightarrow{p} N_\ell$、$\ell = 1$、2、3。其余结论随即得出。定理得证。

□

定理4的证明。在条件 (11) 下，根据引理 1，我们有 $T_n = \tilde{T}_n + \mathcal{O}(n^{-1/2})$，因此$T_n/n = \tilde{T}_n/n + \mathcal{O}(n^{-3/2})$。由此可得 $\mathrm{E}(T_n/n) = \mathrm{E}(\tilde{T}_n/n) + \mathcal{O}(n^{-3/2}) =$ 希尔伯特-施密特独立性准则 $+ \mathcal{O}(n^{-3/2})$。因此，$\sqrt{n}(T_n/n -$ 希尔伯特-施密特独立性准则$) = \sqrt{n}(T_n/n -$ 希尔伯特-施密特独立性准则$) + \mathcal{O}(n^{-1})$。现在

$$\begin{aligned}\tilde{T}_n/n &= n^{-2}\sum_{i=1}^{n}[\tilde{K}(x_i,x_i)\tilde{L}(y_i,y_i)] + \tfrac{n-1}{n}\binom{n}{2}^{-1}\sum_{1\leq i<j\leq n}[\tilde{K}(x_i,x_j)\tilde{L}(y_i,y_j)] \\ &= \binom{n}{2}^{-1}\sum_{1\leq i<j\leq n}[\tilde{K}(x_i,x_j)\tilde{L}(y_i,y_j)] + \mathcal{O}(n^{-1}).\end{aligned}$$

˜ ˜令 (x', y') 为 (x, y) 的一个独立副本。由于 $\mathrm{E}[K(x,x')L(y,y')] =$ 希尔伯特-施密特独立性准则 > 0，x 且 y 不独立。因此，根据 Serfling (1980, 定理 A，第 5.5.1 节)，有 $\mathrm{E}[\tilde{K}(x,x')\tilde{L}(y,y')|(x',y')] \neq 0$，且 Var $\sigma^2 = \mathrm{E}[\tilde{K}(x,x')\tilde{L}(y,y')|(x,y)]\Big\} > 0$；当 $n \to \infty$ 时，我们有 $\sqrt{n}\left(\tilde{T}_n/n - \text{希尔伯特-施密特独立性准则}\right) \xrightarrow{\mathcal{L}} N(0, 4\sigma^2)$。由此可得 $\sqrt{n}\,(T_n/n -$ 希尔伯特-施密特独立性准则$) \xrightarrow{\mathcal{L}} N(0, 4\sigma^2)$。在条件 (11) 下，根据定理 3，我们有 $\hat{}\beta_0 \xrightarrow{p} \beta_0$、$\hat{}\beta_1 \xrightarrow{p} \beta_1$ 和 $\hat{}d \xrightarrow{p} d$。令 $\Phi(\cdot)$ 表示标准正态分布的累积分布函数。因此，在条件 (11) 和局部备择假设 (20) 下，当 n 较大时，对于任意显著性水平 α，有

$$\begin{aligned}Pr\left[T_n \geq \hat{\beta}_0 + \hat{\beta}_1\chi_d^2(\alpha)\right] &= Pr\left[\tfrac{\sqrt{n}(T_n/n - \mathrm{HSIC})}{2\sigma} \geq \tfrac{\hat{\beta}_0 + \hat{\beta}_1\chi_d^2(\alpha)}{2\sqrt{n}\sigma} - \tfrac{\sqrt{n}\mathrm{HSIC}}{2\sigma}\right] \\ &= \Phi\left[-\tfrac{\beta_0 + \beta_1\chi_d^2(\alpha)}{2\sqrt{n}\sigma} + \tfrac{n^{\Delta}h}{2\sigma}\right][1 + o(1)] \\ &= \Phi\left[n^{\Delta}h/(2\sigma)\right][1 + o(1)],\end{aligned}$$

当 $n \to \infty$ 时趋近于 1。定理得证。 □

参考文献

Dubin, J.A. 和 Müller, H.-G.（2005）。多元纵向数据的动态相关。美国统计学会会刊，100(471)：872–881。

Fukumizu, K.、Bach, F. R. 和 Jordan, M. I.（2004）。利用再生核希尔伯特空间进行监督学习的降维。机器学习研究期刊，5(Jan)：73–99。

Górecki, T. 和 Smaga, Ł.（2017）。函数型数据的多变量方差分析。应用统计学杂志，44(12)：2172–2189。

Gretton, A.、Bousquet, O.、Smola, A. 和 Schölkopf, B.（2005a）。使用希尔伯特–施密特范数度量统计依赖关系。载于国际算法学习理论会议，第 63–77 页。Springer。Gretton, A.、Fukumizu, K.、Teo, C.、Song, L.、Schölkopf, B. 和 Smola, A.（2007）。核独立性统计检验。载于 Platt, J.、Koller, D.、Singer, Y. 和 Roweis, S.（编），神经信息处理系统进展，第 20 卷。Curran Associates, Inc.
Gretton, A.、Herbrich, R.、Smola, A.、Bousquet, O. 和 Schölkopf, B.（2005b）。用于度量独立性的核方法。机器学习研究期刊，6(70)：2075–2129。Hastie, T.、Tibshirani, R. 和 Hastie, M. T.（2022）。软件包“ProDenICA”。Kosorok, M. R.（2009）。对“布朗距离协方差”的讨论。应用统计年鉴，3(4)：1270–1278。Miao, R.、Zhang, X. 和 Wong, R. K.（2022）。用于函数型数据的基于小波的独立性检验及其在脑磁图功能连接中的应用。美国统计学会会刊，第 1–14 页。Pearson, K.（1895）。关于双亲情况下回归与遗传的注记，《伦敦皇家学会会刊》，58，240–242。*K Pearson*。Pfister, N. 和 Peters, J.（2017）。dhsic：通过希尔伯特–施密特独立性准则进行独立性检验。*R*软件包版本，2。Ramsay, J. O. 和 Silverman, B. W.（2005）。函数型数据分析。Springer，纽约，第 2 版。Serfling, R. J.（1980）。数理统计近似定理。Wiley，纽约。Shen, C.、Priebe, C. E. 和 Vogelstein, J. T.（2019）。从距离相关到多尺度图相关。美国统计学会会刊。Smola, A.、Gretton, A.、Song, L. 和 Schölkopf, B.（2007）。分布的希尔伯特空间嵌入。载于国际算法学习理论会议论文集，4754：13–31。Székely, G. J.、Rizzo, M. L. 和 Bakirov, N. K.（2007）。通过距离相关度量和检验依赖关系。统计年鉴，35(6)：2769–2794。Zhang, J.-T.（2005）。卡方型混合分布的近似与渐近分布及其应用。美国统计学会会刊，100(469)：273–285。Zhang, J.-T.（2011）。不等单元样本量和不等单元协方差矩阵的双因素多元方差分析。*Technometrics*，53(4)：426–439。Zhang, J.-T.（2013）。函数型数据方差分析。CRC出版社。

Zhang, J.-T., Guo, J. 和 Zhou, B.（2022）。可分度量空间中多个分布的相等性检验：一种基于最大均值差异的方法。《计量经济学杂志》。Zhang, K., Peters, J., Janzing, D. 和 Schölkopf, 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*UAI*。Zhang, Q., Filippi, S., Gretton 等（2018）。用于独立性检验的大规模核方法。《统计计算》。Zhou, Y., Lin, S.-C. 和 Wang, J.-L.（2018）。纵向数据的局部和全局时间相关性。《多元分析杂志》，167：1–14。Zhu, T., Zhang, J.-T. 和 Cheng, M.-Y.（2022）。通过 Lawley–Hotelling 迹检验对函数型数据进行单因素多元方差分析。《多元分析杂志》，192：105095。