拓扑因果效应

KwanghoKim*，**Hajin Lee**统计系，韩国高丽大学，韩国首尔城北区安岩路145号，邮编02841 {kwanghk, hlee2745}@korea.ac.kr

arXiv:2603.02289v1 [stat.ME] 2 Mar 2026

摘要

当结果产生于复杂的非欧几里得空间时，因果效应的估计尤其具有挑战性，因为传统方法往往无法捕捉有意义的结构变异。我们提出了一个用于拓扑因果推断的框架，通过潜在结果拓扑结构的差异来定义处理效应，并使用持久图的幂加权轮廓函数对其进行概括。我们在完全非参数模型中提出了一种高效的双重稳健估计量，证明了函数型弱收敛，并构造了一个用于检验无拓扑效应这一原假设的正式检验。实证研究表明，所提方法能够可靠地量化各种复杂结果类型中的拓扑处理效应。

1 引言

因果推断已成为许多学科的核心工具，为研究干预效应提供了超越单纯关联的统计框架。在潜在结果框架（Imbens & Rubin, 2015）中，因果效应通过比较反事实结果来定义，即在不同处理分配下本会发生什么。然而，随着现代科学研究的结果日益复杂，依赖欧几里得摘要的标准因果估计量和方法可能无法检测到结果底层结构中由干预引发的有意义变化。尽管这类问题具有重要的现实意义，但针对本质上非欧几里得、高维且无结构结果的因果推断，方法学研究相对较少。

在这项工作中，我们关注这样一类情形：具有科学意义的干预效应体现为结果拓扑结构的变化，而非简单欧几里得摘要的偏移。这类效应自然出现在生物医学领域，例如处理会改变大分子构象或折叠模式（Kovacev-Nikolic 等，2016；Cang 和 Wei，2018；Axelrod & Gomez-Bombarelli，2022）；也出现在神经科学领域，例如刺激会重塑脑连接（Sizemore 等，2019）；还出现在信号处理和医学影像领域，在这些领域中，研究目标是检测动力系统或诊断图像中的结构变化（Kim 等，2018；Gholizadeh 和 Zadrozny，2018）。

为形式化并估计此类效应，我们引入一类因果估计目标，利用拓扑数据分析（TDA）工具，量化干预引起的复杂结果拓扑结构变化。TDA 从复杂数据中提取稳健的、多尺度的几何与拓扑描述符（Carlsson，2020）。我们以持久同调为基础；它概括了随着分辨率参数变化，拓扑特征（例如连通分支、环和空腔）如何出现与消失（Chazal 和 Michel，2021）。尽管 TDA 已被用于提升分布偏移和扰动条件下的预测稳健性（例如 Carrière 等，2020；2021；Kim 等，2020），并已作为正则化项纳入传统的平均处理效应分析流程（Farzam 等，2025），但据我们所知，尚无现有框架能够（i）直接以拓扑摘要定义因果估计目标；并且（ii）提供相应的非参数估计与推断。我们的工作填补了这一空白：定义一种具有拓扑意识的因果估计目标，并开发相应的高效双重稳健估计量，且可进行有效推断。

示例说明。我们使用一个大分子数据集（Axel-rod & Gomez-Bombarelli，2022）来展示所提出的方法。该数据集见图1，并将在第6节再次讨论。每个分子都

*通讯作者。两位作者对本研究的贡献相同。

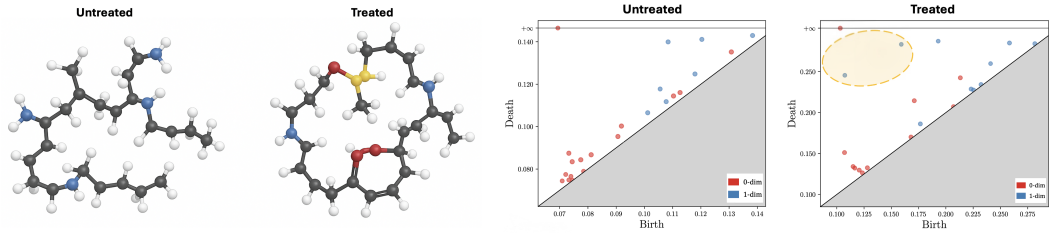


图1：左：未处理与已处理大分子结构示例。右：对应的持久图，突出显示处理诱导的一阶同调特征变化。

表示为一个简化的图，不同样本中的图大小各异。我们考虑一种假设的化学处理，它会使分子几何中出现更多类似环的结构。使用传统的欧几里得摘要可能难以检测此类变化，但在一维持久同调中，这些变化十分明显，表现为相应持久图的偏移。随后，我们将这些持久图转换为便于估计和推断的函数型摘要（Chazal 等人，2014；Bubenik 等人，2015）。这一视角将我们的方法与近期关于函数型结果因果推断的研究联系起来（Ecker 等人，2024；Testa 等，2025）。这一研究方向的方法论发展仍处于早期阶段，尤其是在全非参数估计和推断方面。

贡献。 我们提出一类新的拓扑因果效应，通过干预引起的持久同调摘要变化来定义；具体而言，它是潜在结果下幂加权轮廓函数的期望对比。该目标旨在捕捉标量或欧几里得摘要无法呈现的结构效应，同时保留拓扑描述符在扰动下的稳定性。我们针对由此得到的函数型目标估计量，开发了一种高效、双重稳健且完全非参数的估计量；在干扰误差满足标准乘积速率条件时，获得快速的 $\sqrt{n}$ 速率，并证明其弱收敛，从而能够在完全非参数的设定下进行有效且易于实施的推断。我们还推导出持久图受到 Wasserstein 扰动时加权轮廓的新的稳定性界，并将这些界与弱收敛结果相结合，构造了针对无拓扑效应原假设的正式检验，该检验具有渐近有效的检验大小和一致性。总之，我们的方法拓展了因果推断的适用范围，使复杂系统中的结构效应能够得到严格分析，并支持在多种现代数据模态中开展因果研究，实证研究也展示了这一点。

2 预备知识：机器学习的拓扑工具

本节简要概述拓扑数据分析中的基础概念，并介绍本文中使用的符号。更多详情请参见 Hatcher（2002）；Edelsbrunner 和 Harer（2010）；Chazal 等（2016），以及附录A。

单纯复形。 为了表示有限样本中复杂数据的形状，我们用单纯复形近似其底层对象；单纯复形是图在更高维度上的推广。给定顶点集 V，一个（抽象）单纯复形是由 V 的有限子集组成的集合 K，满足：若 $\sigma \in K$ 且 $\tau \subseteq \sigma$，则 $\tau \in K$。每个 $\sigma \in K$ 都称为一个单形，其维数为 $\dim(\sigma) = |\sigma| - 1$，而 K 的维数是其单纯形中的最大单形维数。特别地，一维单纯复形对应于一个图（可能含有孤立顶点）。在实际应用中，单纯复形通常根据观测到的点云 ˇ，采用标准规则构造，例如 Vietoris–Rips复形或 Čech复形。

持久同调与持久图。 滤过是一族嵌套的单纯复形 $\mathcal{F} = \{K(t) : t \in \mathbb{R}\}$，满足 $t \le t'$ 蕴含 $K(t) \subseteq K(t')$。滤过通常由单调过滤函数 $f: K \to \mathbb{R}$ 生成，该函数满足：只要 $\tau \subseteq \sigma$，就有 $f(\tau) \le f(\sigma)$；并令 $K(t) = f^{-1}((-\infty, t])$。持久同调分别针对每个同调维数 $d = 0, 1, 2, \ldots\ldots$，概括拓扑特征如何随着过滤参数 t 增大而出现和消失（例如，$d = 0$ 连通分支、$d = 1$ 环、$d = 2$ 空腔）。我们将 $\mathcal{H}_d\{K(t)\}$记作

d 阶同调群在水平 t处的同调群，并用其出生时间 a 和死亡时间 $b > a$表示每个特征；由此得到的出生–死亡对 $(a,\ b)$ 构成持久图。形式上，由 $\mathcal{F}$ 诱导的 d 阶持久图记为 $\mathcal{D}_d(\mathcal{F})$，它是以下空间中的一个多重点集

$$\mathbb{R}^{2+} := \{(a,b) \in (\mathbb{R} \cup \{\infty\})^2 : a < b\}.$$

当滤过可由上下文明确判断时，我们将 $\mathcal{D}_d$ 记作 $\mathcal{D}_d(\mathcal{F})$，并用 $\mathcal{D}$ 表示一个泛指的持久图。

加权轮廓。 由于 $\mathcal{D}$ 是一个多重集，将其嵌入函数空间通常会很方便。给定 $\mathcal{D}$，对每个 $p = (a_p,\ b_p) \in \mathcal{D}$，定义帐篷函数 $\Lambda_p \colon \mathbb{R} \to \mathbb{R}$,

$$\Lambda_p(t) = \max\{0, \min\{t - a_p, b_p - t\}\}, \qquad t \in \mathbb{T}, \tag{1}$$

其中 $\mathbb{T} \subset \mathbb{R}$ 是一个固定的紧区间。加权轮廓是这些帐篷函数的归一化加权平均,

$$\phi(t; \mathcal{D}) = \frac{\sum_{p \in \mathcal{D}} w_p\, \Lambda_p(t)}{\sum_{p \in \mathcal{D}} w_p}, \qquad t \in \mathbb{T},$$

其中 $w_p \geq 0$ 是与点 p对应的权重，且 $\sum_{p \in \mathcal{D}} w_p > 0$。一种常见的选择是幂权重 $w_p = (b_p - a_p)^r$，其中 $r > 0$，由此得到幂加权轮廓

$$\phi(t; \mathcal{D}, r) = \frac{\sum_{p \in \mathcal{D}} (b_p - a_p)^r\, \Lambda_p(t)}{\sum_{p \in \mathcal{D}} (b_p - a_p)^r}, \qquad t \in \mathbb{T}. \tag{2}$$

较大的 r 会更加突出寿命较长的特征，而较小的 r 则会相对增加寿命较短特征的权重。由于 r 仅控制突出哪些特征，不影响识别、估计或推断，因此在其固定时，我们在记号中省略 r。

下面的引理记录了轮廓函数的一项基本正则性性质，我们将在第 5 节中用到。**引理 2.1**（加权轮廓的利普希茨稳定性）。 对于任意 $\delta > 0$,

$$\sup_{|s-t| \leq \delta} \big|\phi(s; \mathcal{D}) - \phi(t; \mathcal{D})\big| \leq \delta,$$

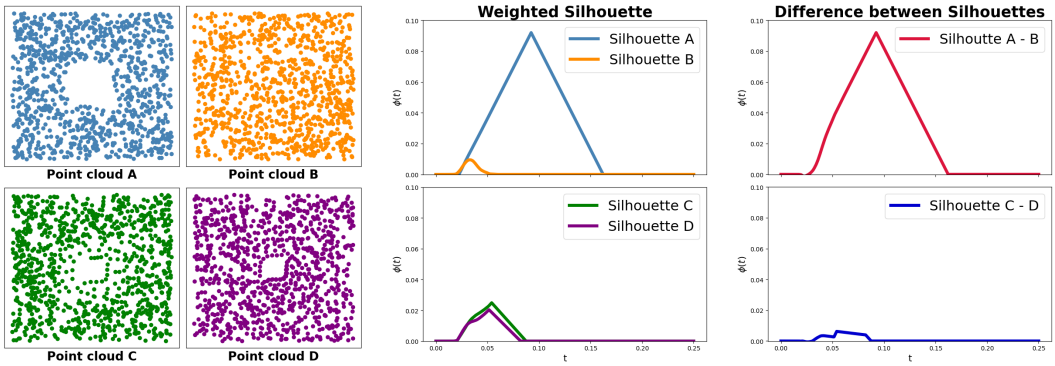
因此

$$\mathbb{E}\Big[\sup_{|s-t| \leq \delta} \big|\phi(s; \mathcal{D}) - \phi(t; \mathcal{D})\big|\Big] \leq \delta.$$

滤过的选择。 计算持久同调描述符需要指定一个滤过$\mathcal{F} = \{K(t) :\ t \in \mathbb{R}\}$，通常的做法是选择一个单纯复形（或立方复形） K，以及一个单调过滤函数 $f \colon K \to \mathbb{R}$，并令 $K(t) = f^{-1}((-\infty,\ t])$。适当的构造取决于数据模态以及希望探究的科学尺度概念。对于点云数据，维托里斯–里普斯滤过和 α-滤过是常见选择；对于数字图像等网格结构数据，在立方复形上使用次水平集或超水平集滤过很自然。对于图数据，可以使用团滤过，或持久同调变换（PHT）（Turner et al., 2014）。这些标准选项使我们的框架能够通过为每种模态配备适当的滤过，适用于广泛的复杂结果类型。若干标准构造（维托里斯–里普斯、 α-和立方滤过）的具体定义汇总于附录A。

3 框架

令 $\{Z_i = (X_i, A_i, Y_i)\}_{i=1}^n$ 表示一个独立同分布观测样本。这里， $A_i \in \mathcal{A} = \{0, 1\}$ 是二元处理（干预）指示变量， $X_i \in \mathcal{X} \subseteq \mathbb{R}^l$ 是一个 l 维协变量向量。令 $\mathcal{F}_i$ 表示由 Y_i 构造的单纯复形滤过，并令 $\mathcal{F}_i^a$ 表示相应的潜在（反事实）滤过，即在处理 $A_i = a$ 下由潜在结果 Y_i^a 构造的滤过。对于每个同调维数 d，将观测持久图和潜在持久图分别定义为 $\mathcal{D}_{i,d} := \mathcal{D}_d(\mathcal{F}_i)$ 和 $\mathcal{D}_{i,d}^a := \mathcal{D}_d(\mathcal{F}_i^a)$。此外，令 $\phi_{i,d}(t) := \phi\{t; \mathcal{D}_{i,d}\}$ 和 $\phi_{i,d}^a(t) := \phi\{t; \mathcal{D}_{i,d}^a\}$（其中 $t \in \mathbb{T}$），表示方程 2 中相应的幂加权轮廓函数，其中 $\mathbb{T}$ 是 ϕ 的定义域。



(a) (b) (c)

图2：揭示 ORBIT 数据集（Adams等人，2017）中一维特征差异的轮廓函数。(a) 点云 A，B，C，D；(b) 其幂加权轮廓；(c) 轮廓对比 $\phi_A - \phi_B$（上）和 $\phi_C - \phi_D$（下）。$\phi_A - \phi_B$中的强信号表明，相对于 B，A 中出现了新的一维特征，而接近于零的 $\phi_C - \phi_D$ 表明 C 与 D之间几乎没有变化。

我们感兴趣的目标参数是拓扑平均处理效应（TATE）。对于每个固定的 d，定义函数型因果效应

$$\psi_d(t) \coloneqq \mathbb{E}\left\{\phi^1_{i,d}(t) - \phi^0_{i,d}(t)\right\}, \qquad t \in \mathbb{T}. \tag{3}$$

等价地，对于选定的最大同调度 $d_{\max}$，我们可以将这些分量堆叠成 $\mathbb{R}^{d_{\max}+1}$值曲线 $\psi(t) = (\psi_0(t) \ldots, \psi_{d_{\max}}(t))$，$\forall t \in \mathbb{T}$，使得 $\psi_d(t)$ 是 $\psi(t)$ 的第 d 个坐标。每个函数 $\psi_d(t)$ 刻画了在过滤尺度 t 下，d维拓扑特征因处理而产生的变化：$\psi_d(t) > 0$ 表明，平均而言，处理组个体在尺度 t 下表现出更多或更强的 d 维特征，而 $\psi_d(t) < 0$ 表明处理会导致其减弱。因此，映射 $t \mapsto \psi_d(t)$ 是希尔伯特空间中的函数型因果效应，描绘了处理如何随着过滤尺度的变化而改变拓扑结构。

将方程 3 解读为处理诱导的拓扑效应，具有若干显著优势。首先，$\psi_d(t)$ 本身具有尺度感知能力：轮廓函数以过滤参数 t 为索引，因此保留了不同几何尺度上的拓扑信息，从而使 $\psi_d(t)$ 能够定位不同几何尺度上的处理诱导拓扑变化。其次，它具有抗噪性：幂加权轮廓会降低短暂且可能是虚假特征的权重，因此 $\psi_d(t)$ 着重体现更可能具有实际意义的持久结构差异。第三，与原始持久图不同，$\psi_d(t)$ 可以向量化，也就是说，它存在于可分希尔伯特空间中，这有助于开展理论分析，并将其整合到基于梯度的机器学习流程中。第四，$\psi_d(t)$ 与函数型因果推断的最新进展自然契合。总的来说，$\psi_d(t)$ 可视为平均处理效应的拓扑感知类比，能够捕捉干预如何重塑不同处理组中底层结果的 d 维同调结构。

一个局限是，轮廓函数可能会掩盖发生变化的同调特征的确切数量，因为它们将多个帐篷函数聚合为一个加权汇总。不过，原则上也可以利用所选层级以内各同调维数上的单个持久景观函数来定义拓扑平均处理效应（TATE），从而可能获得更精细的拓扑分辨率。尽管如此，用单条轮廓曲线表示每个图，能够简洁且易于解读地概括给定同调维数内的整体拓扑变化；而幂参数 r 则提供了一种直观的方式，可在需要时突出显示最持久的特征。

尽管可以先将每个持久图向量化为一个临时的欧几里得特征向量，再应用标准的平均处理效应估计量，但这类启发式方法通常缺乏与底层拓扑对象之间有原则性的联系。我们的方法有着根本区别：直接以持久图几何来定义因果估计目标，并将轮廓函数用作函数嵌入，从而在希尔伯特空间中得到适定的目标。这带来了连贯的拓扑因果效应概念，以及与该目标明确对齐的双重稳健估计和有效推断。

图2展示了目标参数（方程3）如何捕捉结构变化。图2a中的点云对（A， B）和（C， D）分别表示处理和对照下的潜在结果。相对于 B，处理在 A 中诱发了一个环，使图2c中出现显著的轮廓对比；而（C， D）的对比接近于零，表明结构变化可以忽略不计。轮廓差异曲线的符号表示拓扑变化的方向：正值对应于处理下出现的特征，负值则对应于相对于对照消失的特征。

为进行识别，我们采用标准的因果假设：(C1) 一致性， $\mathcal{F}_i = \mathcal{F}_i^a$当且仅当$A_i = a$；(C2)无未测量混杂，$A_i \perp\!\!\!\perp \phi_{i,\,d}^{a}(t) \mid X_i$；以及 (C3) 正值性， $\mathbb{P}(A_i = a \mid X_i) > 0$ 几乎处处成立，对于所有 $t \in \mathbb{T}$、$a \in \mathcal{A}$ 以及同调维数 d 均如此。尽管我们的目标参数更为复杂，但这些识别条件与标准因果推断设定中的条件基本平行（例如 Imbens & Rubin，2015）。我们假设本文始终满足假设 (C1)–(C3)。需要注意的是，分量级识别足以在每个过滤尺度 t 上识别函数效应 $\psi_d(t)$。于是，对于每个 d，目标 $\psi_d(t)$ 可识别为

$$\psi_d(t) = \mathbb{E}[\mathbb{E}\{\phi_{i,d}(t) \mid X_i, A_i = 1\} - \mathbb{E}\{\phi_{i,d}(t) \mid X_i, A_i = 0\}] \tag{4}$$

$$= \mathbb{E}\left\{ \frac{A_i\, \phi_{i,d}(t)}{\pi(X_i)} - \frac{(1 - A_i)\, \phi_{i,d}(t)}{1 - \pi(X_i)} \right\}. \tag{5}$$

上述识别表达式促使我们采用插件估计和逆概率加权估计量，这些内容将在下一节中作更详细的讨论。

备注1 （r的选择）。 拓扑平均处理效应 $\psi_d(t)$ 取决于幂参数 r，该参数控制轮廓函数对长寿命与短寿命持久性对的强调程度。我们将 r 视为特定问题的调优参数：实践者可以利用领域知识来选择它（例如，使用较大的r 来提高持久特征的权重），也可以采用简单的数据驱动准则，例如检查 ψ_d 在不同候选 r 值下的$\widehat{\ }$稳定性（例如，在某个处理组内进行验证步骤，或通过交叉验证）。在我们的实验中，在适度范围的 r 值内，估计效应的定性形态和统计显著性保持稳定，这表明结果对其具体取值不太敏感。

4 估计

本节将针对方程3中定义的 ψ_d 提出一种估计策略。下文中，只要不会产生歧义，我们便省略标记第 i个观测的下标，并用一般随机变量代替其样本实现值。为方便起见，将倾向评分和条件轮廓回归函数定义为

$$\pi(x) \coloneqq \mathbb{P}(A = 1 \mid X = x), \qquad \mu_a(t, x; d) \coloneqq \mathbb{E}\{\phi_d(t) \mid X = x, A = a\}.$$

$\widehat{\ }$令$\hat{}\pi$ 和$\hat{}\mu_a$ 分别表示 π 和 μ_a的估计量。我们用 $\mathbb{P}$ 表示用于拟合干扰参数估计量（$\hat{}\pi$，$\hat{}\mu_0$，$\hat{}\mu_1$）的经验分布，并假设可以获得一个独立于 $\mathbb{P}$的单独经验$\widehat{\ }$分布 $\mathbb{P}_n$，用于构造最终估计量。

受方程 4 和方程 5 中识别表达式的启发，我们提出插件估计（PI）回归估计量和逆概率加权（IPW）估计量，形式如下：

$$\widehat{\psi}_{\mathrm{PI},d}(t) = \mathbb{P}_n\big\{\hat{\mu}_1(t, X; d) - \hat{\mu}_0(t, X; d)\big\}, \tag{6}$$

$$\widehat{\psi}_{\mathrm{IPW},d}(t) = \mathbb{P}_n \left\{ \frac{A\, \phi_d(t)}{\hat{\pi}(X)} - \frac{(1 - A)\, \phi_d(t)}{1 - \hat{\pi}(X)} \right\}. \tag{7}$$

$\widehat{\ }$ $\widehat{\ }$估计量 $\psi_{\mathrm{PI},d}$ 和$\psi_{\mathrm{IPW},d}$ 分别继承了$\hat{}\mu_a$ 和$\hat{}\pi$的收敛速率。在许多观察性研究中，逆概率加权估计量颇具吸引力，因为对处理分配机制建模通常比对函数型$\widehat{\ }$结局的完整条件均值建模更可行（Imbens & Rubin，2015）。在我们的设定中，$\psi_{\mathrm{IPW},d}$ 可能尤为方便，因为估计函数回归 $\mu_a(\cdot, \cdot; d)$ 可能远比估计标量倾向评分 π更具挑战性。

可以利用半参数效率理论构造一个效率更高的估计量。定义未中心化的有效影响函数（EIF）如下：

$$\varphi_d(t, Z; \eta) \coloneqq \mu_1(t, X; d) - \mu_0(t, X; d) + \left\{ \frac{A}{\pi(X)} - \frac{1-A}{1-\pi(X)} \right\} \{\phi_d(t) - \mu_A(t, X; d)\}, \quad (8)$$

whe收集干扰函数。根据构造， $\mathbb{E}\{\varphi_d(t, Z;\ \eta)\} = \psi_d(t)$ fo r eac h $t \in \mathbb{T}$。影响函数表示得出一个增广逆概率加权估计量，

$$\widehat{\psi}_{\mathrm{AIPW},d}(t) \coloneqq \mathbb{P}_n\{\varphi_d(t, Z; \widehat{\eta})\} \equiv \mathbb{P}_n\{\widehat{\varphi}_d(t)\}, \quad (9)$$

$\widehat{\ }$ $\widehat{\ }$其中， $\widehat{\eta} = \eta(\mathbb{P}) = \{\hat{\pi}, \hat{\mu}_0, \hat{\mu}_1\}$ 使用干扰函数训练样本 $\mathbb{P}$进行估计。这种构造产生了通常的二阶余项结构和熟悉的双重稳健性质，使得 $\psi_{\mathrm{AIPW},d}$ 在对干扰参数估计量施加较弱非参数条件时能够达到快速收敛率（ Kennedy，2016 ；2024）。

建立增广逆概率加权估计量渐近性质有两种标准方法。一种方法是对影响函数指标函数类 $\{\varphi_d(t, \cdot; \eta) : t \in \mathbb{T}\}$施加经验过程条件（例如 Donsker 条件或适当的熵条件），并对干扰参数估计量施加相应的正则性条件，但对于现代高度自适应的干扰函数学习而言，这些条件可能过于严格。另一种方法是样本分割，即将干扰函数拟合与最终去偏步骤分开，从而允许使用任意复杂的干扰参数估计量，并减轻对过拟合的担忧。原则上，我们不限定干扰函数的学习方式。不过，为了表述清晰，我们在全文采用样本分割设定：在 $\mathbb{P}$ 上拟合干扰参数估计量，并在一个独立样本 $\mathbb{P}_n$ 上构造最终估计量。（见假设（A2）。）

$\widehat{\ }$**注释 2 (**样本分割)。 通过随机数据分割可以获得独立样本 $\mathbb{P}_n$ 和 $\mathbb{P}$ 。通过交叉拟合可以恢复全样本效率（例如，*Chernozhukov* 等，*2018*；*Newey* 和 *Robins*，*2018*）。为使阐述清晰，我们采用单次分割，但扩展到多次分割也很简单（例如，*Kennedy*，*2019*；*2023*；*Kennedy* 等，*2023*）。

5 推断

对于标量结局的平均处理效应估计，保证理想大样本性质（例如 $\sqrt{n}$ 阶收敛和渐近正态性）的条件已广为人知（例如，Kennedy，2016；2024）。在我们的设定中，结局是一条以过滤尺度为索引的曲线，因此，有效推断需要额外的结构来控制所得随机过程的行为。

我们首先考虑一个简化设定，其中处理分配机制完全已知，如同随机试验中那样。具体而言，我们假设真实倾向评分 π 几乎必然已知。在这种情况下，公式 7 中的逆概率加权估计量是无偏的。

然而，若要使整个估计效应曲线在 $\ell^\infty(\mathbb{T})$ 中弱收敛，仅假设逐点二阶矩并不足够；还需要指标经验过程 $\{\xi_d(\cdot,\ t) :\ t \in \mathbb{T}\}$ 满足函数型中心极限定理。例如，可以假设指标函数类是 $\mathbb{P}$-Donsker 类（或满足带有适当可测性的标准熵与包络条件），从而保证这一点。在我们的设定中，这些条件是合理的，因为轮廓函数 $\phi_d(t)$ 关于指标 t 是利普希茨连续的（参见附录C.2中的备注3）。以下定理表明，当 π 已知时，$\sim\psi_{\mathrm{IPW},\ d}$ 在温和条件下弱收敛到一个高斯过程，该过程取值于 $\ell^\infty(\mathbb{T})$。

定理 5.1。 对于 $t \in \mathbb{T}$，定义 $\xi_d(Z, t)$: $= \left\{ \frac{A}{\pi\ X()} - \frac{1-A}{1-\pi\ X()} \right\} \phi_d(t)$，其中 π 已知。假设 $\sup^{t\in\mathbb{T}} \mathit{var}(\xi_d(Z, t)) < \infty$，并且函数类 $\{\xi_d(\cdot, t)\colon\ t \in \mathbb{T}\}$ 是 $\mathbb{P}$-*Donsker* 类（或满足在 $\ell^\infty(\mathbb{T})$ 中成立的函数型中心极限定理的任何其他标准充分条件）。那么

$$\sqrt{n}\{\widehat{\psi}_{\mathrm{IPW},d}(t) - \psi_d(t)\} \rightsquigarrow \mathbb{G}_d(t) \quad \mathit{in} \quad \ell^\infty(\mathbb{T}).$$

其中， $\mathbb{G}_d$ 是均值为零的高斯过程，其协方差函数为 $\mathit{cov}\{\mathbb{G}_d(s)\ \mathbb{G}_d(t)\} = \mathit{cov}\big(\xi_d(Z, s)\ \xi_d(Z, t)\big), \forall s, t \in \mathbb{T}$.

在观察性研究场景中，通常最好使用公式9中定义的增广逆概率加权估计量。下文中，我们使用记号 $\|\cdot\|_{\mathbb{P},\ q}$ 表示 $L_q(\mathbb{P})$-范数。为分析渐近性质，我们引入以下假设。

(A1) $\|1/\widehat{\pi}\|_\infty < \infty$ 和 $\|1/(1-\widehat{\pi})\|_\infty < \infty$。

(A2) 样本分割：讨厌参数估计 $\widehat{\eta}$ 在 $\mathbb{P}$ 上拟合，最终估计量在一个独立样本 $\mathbb{P}_n$（$\mathbb{P}_n \perp\!\!\!\perp \mathbb{P}$）上计算。

(A3) 截面相合性：$\|\widehat{\varphi}(t) - \varphi(t)\|_{\mathbb{P},2} = o_{\mathbb{P}}(1)$，$\forall t \in \mathbb{T}$。

(A4) 讨厌参数估计的速率条件：$\|\widehat{\pi}(X) - \pi(X)\|_{\mathbb{P},2} \left\{ \sum_{a\in\mathcal{A}} \|\widehat{\mu}_a(t, X;\, d) - \mu_a(t, X;\, d)\|_{\mathbb{P},2} \right\} = o_{\mathbb{P}}(n^{-1/2})$，$\forall t \in \mathbb{T}$。

(A5) 一致一致性：对所有 $x \in \mathcal{X}$，$\mu_a(t,\ x;\ d)$ 关于 t 在 $\mathbb{T}$ 上一致利普希茨，且下列条件成立：

$$\sup_{t\in\mathcal{T},x\in\mathcal{X}} \left|\widehat{\mu}_a(t,x;d) - \mu_a(t,x;d)\right| = o_{\mathbb{P}}(1), \quad \sup_{x\in\mathcal{X}} \left|\widehat{\pi}(x) - \pi(x)\right| = o_{\mathbb{P}}(1).$$

(A5′) $\widehat{\mu}_a$ 的利普希茨模：对于 $\delta > 0$ 和 $a \in \mathcal{A}$，$\widehat{\mu}_a(\cdot, X; d)$ 满足 $\mathbb{E}\left[\sup_{|s-t|\le\delta} |\widehat{\mu}_a(s, X;\ d) - \widehat{\mu}_a(t, X;\ d)|\right] \le L_{\hat{\mu}}\delta$，其中 $L_{\hat{\mu}}$ 为某个正常数。

假设 (A1)–(A4) 是半参数因果推断中的标准假设（例如，Kennedy，2019年；Kennedy 等，2023；Kennedy，2024）。（A1）是一个温和的有界性条件，而（A2）将讨厌参数拟合与最终经验平均分开，从而避免对讨厌参数学习器施加经验过程限制。（A3）确保估计的影响函数在每个 t 处满足 $L_2(\mathbb{P})$ 一致性，而（A4）确保二阶冯·米塞斯余项为 $o_{\mathbb{P}}(n^{-1/2})$。对于函数型讨厌参数 $\widehat{\mu}_a(\cdot,\cdot;d)$，我们考虑两种备选的正则性条件。Testa 等人（2025）所采用的假设（A5′）通过要求对 $\sup_{|s-t|\le\delta}|\widehat{\mu}_a(s,X;d) - \widehat{\mu}_a(t,X;d)|$ 施加 L_1 界，从而直接对 $\widehat{\mu}_a$ 的随机样本路径施加利普希茨模。相比之下，假设（A5）仅要求目标回归 $\mu_a(\cdot,\cdot;d)$ 关于 t 一致利普希茨，且 $\widehat{\mu}_a$ 一致收敛于 μ_a。事实上，在诸如 $\mathbb{E}\{\sup_{t\in\mathbb{T}}|\widehat{\mu}_a(t,X;d) - \mu_a(t,X;d)|\} = o(1)$ 这样的温和可积性条件下，我们有

$$\mathbb{E}\Big[\sup_{|s-t|\le\delta}\left|\hat{\mu}_a(s,X;d) - \hat{\mu}_a(t,X;d)\right|\Big] \le L_\mu\,\delta + o(1),$$

因此，只要 $\widehat{\mu}_a$ 一致收敛于一个利普希茨极限，它本身不必是光滑的。

根据$\widehat{\mu}_a$的学习方式，假设（A5）和假设（A5′）中的正则性要求通常可以放宽；如以下例子所示，在某些情况下甚至可以完全避免这些要求。

例*5.1*（函数型线性平滑器）。考虑用于 μ_a 的函数型线性平滑器（例如 *Reiss*等人，*2017*）：假设对于每个 d 和 $a \in \{0,1\}$，拟合回归都可表示为 $\widehat{\mu}_a(t,X;\ d) = \sum_{j=1}^n L_{d,j}(X)\,\phi_{j,d}(t)$，$t \in \mathbb{T}$，其中 $\phi_{j,d}(t) = \phi\{t;\mathcal{D}_{j,d}\}$ 是训练轮廓，$L_{d,j}(X)$ 是平滑器确定的（可能依赖于数据的）权重，并在 X 处取值。那么，对于任意 $\delta > 0$，

$$\begin{aligned}\mathbb{E}\Big[\sup_{|s-t|\le\delta}\left|\hat{\mu}_a(s,X;d) - \hat{\mu}_a(t,X;d)\right|\Big] &\le \mathbb{E}\Big[\sum_{j=1}^n |L_{d,j}(X)| \sup_{|s-t|\le\delta}\left|\phi_{j,d}(s) - \phi_{j,d}(t)\right|\Big] \\ &\le \delta\,\mathbb{E}\Big[\sum_{j=1}^n |L_{d,j}(X)|\Big],\end{aligned}$$

其中，最后一个不等式由引理 *2.1* 得出。特别地，若 $\mathbb{E}\{\sum_{j=1}^n |L_{d,j}(X)|\} < \infty$，则假设（*A5′*）在 $L_{\hat{\mu}} = \mathbb{E}\{\sum_{j=1}^n |L_{d,j}(X)|\}$ 下成立。这一验证仅依赖于平滑权重的矩控制以及轮廓函数内在的利普希茨性质。因此，对于这类估计量，可以不再要求假设（*A5*），并以更简单的原始条件 $\mathbb{E}\{\sum_{j=1}^n |L_{d,j}(X)|\} < \infty$ 取代假设（*A5′*）。普通最小二乘法是一个值得注意的特殊情形。

以下弱收敛结果是我们推断框架的主要技术要素。在附录C.3中，我们在假设（A5′）下证明了该结果。附录C.4随后表明，在假设（A5）下也能得到相同的结论。

定理 5.2。 在假设（*A1*）–（*A4*）以及（*A5′*）或（*A5*）之一成立的条件下，我们有

$$\sqrt{n}\{\widehat{\psi}_{\mathrm{AIPW},d}(t) - \psi_d(t)\} \rightsquigarrow \mathbb{G}_d(t) \quad \textit{in} \quad \ell^\infty(\mathbb{T}),$$

其中 $\mathbb{G}_d$ 是一个零均值高斯过程，其协方差满足 $\mathrm{cov}\{\mathbb{G}_d(s)\ \mathbb{G}_d(t)\} = \mathrm{cov}\{\varphi_d(s, Z;\eta)\ \varphi_d(t, Z;\eta)\}$，对所有 s、$t \in \mathbb{T}$ 成立。

逐点 $\sqrt{n}$-一致性与渐近正态性是定理 5.2 的直接推论。特别地，对每个固定的 $t \in \mathbb{T}$、$\psi_{\mathrm{AIPW},d}(t)$ 达到半参数效率界，其渐近方差为 $\mathrm{var}\{\varphi_d(t, Z;\eta)\}$，并且只要干扰误差满足乘积速率条件（例如，每一项在 $L_2(\mathbb{P})$ 中以 $o_{\mathbb{P}}(n^{-1/4})$ 的速率收敛），即可达到 $\sqrt{n}$ 速率。相比之下，PI 估计量与逆概率加权估计量通常分别由 $\hat{\mu}_a - \mu_a$ 和 $\hat{\pi} - \pi$ 引起一阶偏差项，因此在灵活的非参数学习下，一般需要显著更强的干扰项速率，才能实现 $\sqrt{n}$-一致推断（Kennedy，2024）。在附录 B 中，我们进一步给出了增广逆概率加权估计量、逆概率加权估计量与 PI 估计量在 $\hat{\mu}_a$ 和 $\hat{\pi}$ 相对其真实值存在固定偏离时的显式偏差—方差分析，从而表明增广逆概率加权估计量在干扰项错设下具有双重稳健性。

基于定理 5.2，我们可以在 $\mathbb{T}$ 上构造同时置信带。构造渐近有效的 $1-\alpha$ 置信带有多种方法，具体取决于人们愿意施加的附加假设有多强，或愿意接受多大的计算复杂度。例如，可以采用 Liebl & Reimherr（2023）提出的枢轴方法，或采用 Pini & Vantini（2017）开发的参数自助法。更详细的讨论和比较，请参见 Testa 等人（2025）。

对许多实践者（甚至可能是大多数实践者）而言，首要的推断目标是确定是否存在任何拓扑效应，而持久同调正是对其进行刻画的工具。基于估计得到的向量化拓扑摘要（例如轮廓函数、持久图像或景观函数）的假设检验，通常无法在持久图的度量空间中得到有效推断。相比之下，我们的框架提供了针对“无拓扑效应”这一原假设的正式检验。为此，我们首先建立了加权轮廓的稳定性界；据我们所知，此类界此前尚未见于文献。

定理 5.3。 令 $W_q(\mathcal{D}, \mathcal{D}')$ 表示两个持久图 $\mathcal{D}$ 和 $\mathcal{D}'$ 之间的 *q-Wasserstein* 距离，并令 $m^\star$ 表示相应的最优匹配 W_1。假设

(A6) 对于方程 *2* 中定义的幂加权轮廓函数，其对应的持久图 $\mathcal{D}$ 有界，即对所有 $p = (a_p, b_p) \in \mathcal{D}$, $-\infty < a_p \le b_p < \infty$，均满足该条件。此外，存在一个全局常数 $L > 0$，使得对所有 $p \in \mathcal{D}$ 以及给定的加权指数 r，$\frac{b_p - a_p}{(b_p - a_p)_r} \le L$。$\mathcal{D}'$ 也满足同样的条件。

考虑分别对应于 $\mathcal{D}$ 和 $\mathcal{D}'$ 的加权轮廓函数 ϕ 和 ϕ'。在假设（*A6*）下，

$$\|\phi - \phi'\|_\infty \le \left(1 + 2Lr\, c^{r-1}\right) W_1(\mathcal{D}, \mathcal{D}'),$$

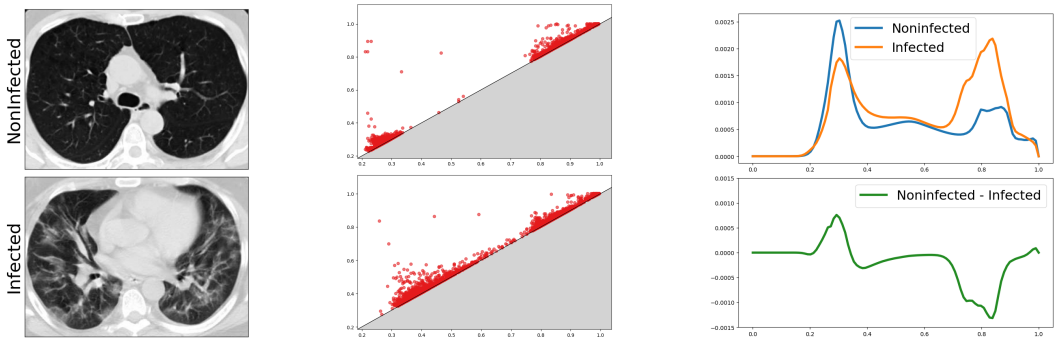
对于某个常数 $c > 0$，该常数仅取决于 r 以及图表 $\mathcal{D}$、$\mathcal{D}'$ 中持久性的上界。

附录 C.5 给出了 Wasserstein 距离、相应最优匹配 $m^\star$ 以及常数 c 的正式定义。基于上述稳定性结果以及我们估计量的高斯弱收敛，据我们所知，我们的框架首次对不存在拓扑效应的原假设进行了正式检验，并且其检验大小渐近正确，且对固定备择具有一致性，如下文正式所述。

推论 5.4. 固定 $d \in \mathbb{N}_0$ 并考虑原假设

$$H_0 : W_1(\mathcal{D}_d^1, \mathcal{D}_d^0) = 0 \quad a.s.,$$

在此条件下，对所有 $t \in \mathbb{T}$，$\psi_d(t) = 0$。令 $T_n := \sqrt{n}\, \|\psi_{\mathrm{AIPW},\, d}\|_\infty$。在假设（*A1*）–（*A6*）下，我们有 $T_n \rightsquigarrow \|\mathbb{G}_d\|_\infty$，其中 $\mathbb{G}_d$ 是定理 *5.2* 中的高斯极限。此外，假设基于估计影响函数的某个高斯或 *Rademacher* 乘子自助过程 $\widehat{\mathbb{G}}_{n,\, d}$ 能一致地估计 $\|\mathbb{G}_d\|_\infty$ 在给定数据条件下的分布。令 $c_{1-\alpha}$ 为条件 $(1-\alpha)$ 分位数 $\|\mathbb{G}_{n,\, d}\|_\infty$。当 $T_n > c_{1-\alpha}$ 时拒绝 H_0 的检验，其渐近检验大小为 α，并且对任何满足 $\|\psi_d\|_\infty > 0$ 的固定备择均具有一致性。



(a) (b) (c)

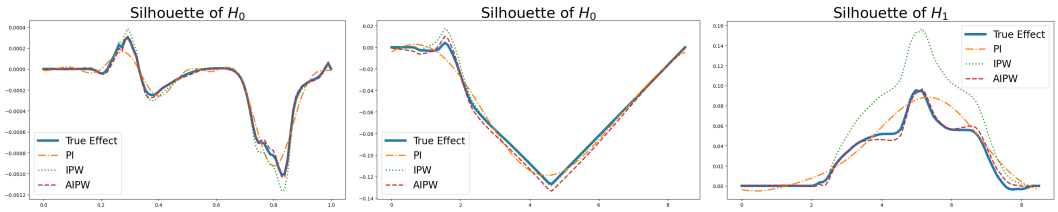
图 3：(a) 未感染（上）和感染（下）患者的 CT扫描图像；(b) 对应的 0 维持久图；(c) 给定 $r = 0.1$ 时未感染和感染患者的平均轮廓函数（上），以及未感染和感染患者平均轮廓函数之差。

在 H_0 下，$T_n = \sqrt{n}\|\psi_{\text{AIPW},\,d}\|_\infty$ 的表现类似于中心化经验过程波动的上确界范数，并收敛到 $\|\mathbb{G}_d\|_\infty$。在任何满足 $\|\psi_d\|_\infty > 0$的固定备择下，T_n发散至 $+\infty$，因此检验的拒绝概率趋于 1。由于$T_n = \sqrt{n}\|\psi_{\text{AIPW},\,d}\|_\infty$ 的极限分布取决于高斯过程 $\mathbb{G}_d$ 的未知分布，我们假设乘子自助法有效，从而可根据数据一致地近似条件 $(1-\alpha)$ 分位数 $\|\mathbb{G}_d\|_\infty$，进而得到渐近正确的临界值。

6 实验

为展示我们方法的有效性，我们在两个半合成数据集和一个合成数据集上开展实验；为完整起见，合成（ORBIT）数据的结果见附录 D.4。对于所有实验，我们构造一个假设数据集（X，A，Y^0，Y^1），其中潜在结果对经过设计，在不同处理组之间呈现出不同的拓扑对比。我们从多元高斯分布生成协变量 $X \in \mathbb{R}^5$，并通过为各子组指定不同的均值向量来构造亚组结构。给定 X，处理 A 以概率 $\pi(X) = expit(-0.5X_1 - 0.1X_2 + 0.6X_3 + 0.1X_4 + 0.1X_5 + 0.5X_2X_3 - 0.7X_1X_3)$ 分配。该处理机制经过设计，使一个子组接受处理的概率高于另一个子组。我们使用基于傅里叶基展开的函数对标量回归来建模轮廓回归函数 μ_a，同时通过随机森林分类器估计倾向评分 π。我们的目标是基于可观测数据估计真实拓扑因果效应。所有实验均重复模拟 20 次。有关实验设置的完整细节，请参见附录 D。

SARS-CoV-2 数据集。 第一个半合成实验使用不同尺寸的图像，并合成协变量和处理分配，同时保留真实结果，从而能够以可控且贴近现实的方式评估因果估计量。我们使用 SARS-CoV-2 数据集（Soares 等人，2020），其中包含从 COVID-19 感染患者和未感染患者身上采集的 CT扫描。感染患者的磨玻璃样阴影和实变发生率较高，这些病灶在 CT扫描中呈现为孤立区域，可通过下星过滤的 0 维持久图捕捉（Iqbal 等人，2025）。图 3a 和 3b 展示了感染者与未感染者的 CT扫描图像差异，以及对应的持久图。图 3c（上）展示了感染患者和未感染患者的平均轮廓函数，其中未感染组在 [0.2, 0.4] 区间的取值较高，而感染组在 [0.7, 0.9] 区间的取值较高。因此，如图 3c（下）所示，若拓扑平均处理效应在区间 $[0.2, 0.4]$（正向）和 $[0.7, 0.9]$（负向）上的绝对值较大，则可认为处理更有效。在本实验中，真实的拓扑平均处理效应是已知的，因为我们通过将 500 个感染样本分配至 Y^0，再与由 75% 未感染样本和 25% 感染样本组成的 Y^1 配对，人工构造了（Y^0，Y^1）。我们根据观测数据计算插件估计、逆概率加权和增广逆概率加权估计量，并将估计结果与真实效应进行比较。



(a) (b) (c)

图 4：真实轮廓函数及其 PI、IPW、AIPW 估计值。(a) 根据 SARS-CoV-2 数据集计算的 0维剪影，(b) 根据 GEOM-Drugs 数据集计算的 0维剪影，以及 (c) 根据 GEOM-Drugs 数据集计算的 1维剪影。

结果。 图 4a 中的蓝色曲线表示真实轮廓，按设计清晰地体现了因果效应。尽管三种估计量都较好地捕捉了真实目标的整体形状，逆概率加权估计量却系统性地高估了真实处理效应，而插件估计量则低估了它。相比之下，增广逆概率加权估计量准确地重建了真实剪影函数，偏差极小，并且与真实值的精确形状高度吻合。

GEOM-Drugs 数据集。接下来，我们通过使用 GEOM-Drugs 数据集（Axelrod & Gomez-Bombarelli，2022）开展半合成实验，在图数据上评估我们的框架。该数据集提供分子化合物的图结构表示。为分析图结构数据，我们首先通过对各图节点的特征进行加权求和，为每个图节点赋予一个标量权重；随后，将每条边的边权重定义为其相邻节点权重的最大值；然后应用子水平集滤过。采用与前一个实验类似的步骤，我们构造 $(Y^0,\ Y^1)$：令 Y^0 包含 1000 个单环图，并将其与 Y^1配对；Y^1由 75% 的双环图和 25% 的单环图混合而成。与前文一样，我们根据观测数据计算 PI、IPW 和 AIPW 估计量，并将估计值与真实效应进行比较。

结果。 三种估计量都与真实的零维轮廓函数高度吻合（图 4b），其呈现负效应，意味着处理使一些连通分支合并。真实的一维轮廓函数（图 4c）显示出显著的正效应，表明处理促使形成新的环，这与我们的实验设计一致。然而，逆概率加权估计量高估了真实的一维处理效应，而 PI 估计量未能捕捉相对复杂的曲率。总体而言，增广逆概率加权估计量对真实轮廓函数给出了最准确、最可靠的近似，这与第 5 节的理论结果一致。

7 讨论

我们提出了因果推断与拓扑数据分析之间一种新颖的联系，从而能够估计因果效应，不仅捕捉均值或方差的变化，还能捕捉结果空间中有意义的内在拓扑结构变化。这拓宽了因果推断的适用范围，并为研究复杂高维情境（包括涉及非结构化数据的情境）中的因果机制开辟了新途径。

若干注意事项与前瞻性解决方案值得关注，它们指明了未来研究富有成效的方向。首先，所提出的拓扑平均处理效应框架旨在捕捉宏观拓扑变化，当主要兴趣在于检测细粒度的局部变化时，其信息量可能较低。在此类情况下，若可行，可并行估计标准因果估计目标，并可能在适当预处理之后进行。与此相关，如第3节所讨论的，轮廓函数并不能精确量化发生变化的同调特征的数量。尽管如此，所提出的估计量以及第4节和第5节中的分析可以自然地扩展到单个持久景观函数，从而在需要时提供更精细的拓扑分辨率。最后，由于使用了持久同调，我们的估计量的构建在计算上可能较为密集。通过采用更高效的拓扑摘要，例如欧拉示性数曲线（Turner et al., 2014），可以减轻这一成本。进一步的扩展包括将该框架应用于更复杂的因果设定，例如连续处理、工具变量设计或纵向暴露。

代码可用性

实现所提出方法的Python代码可在 https://github.com/ kwangho-joshua-kim/top-causal-effect 获取。

致谢

本研究得到韩国国家研究基金会（NRF）由韩国政府（MSIT）资助的项目（编号：RS-2022-NR068754、RS-2024-00335008、RS-2025-24534596）、韩国高丽大学项目（K2304801）以及三星科学技术基金会项目（项目编号：SSTF-BA2502-01）的支持。

参考文献

Henry Adams、Tegan Emerson、Michael Kirby、Rachel Neville、Chris Peterson、Patrick Shipman、Sofya Chepushtanova、Eric Hanson、Francis Motta 和 Lori Ziegelmeier。《持久图像：持久同调的稳定向量表示》。机器学习研究期刊，18(8)：1–35，2017。

Simon Axelrod 和 Rafael Gomez-Bombarelli。Geom：用于性质预测和分子生成的能量标注分子构象。《科学数据》，9(1):185，2022。

Patrick Billingsley。概率测度的收敛。约翰·威立父子出版公司。*INC*，纽约，2(2.4)，1999。

Peter Bubenik等人。使用持久景观进行统计拓扑数据分析。*J. Mach. Learn.Res.*，16(1):77–102，2015。

Zixuan Cang 和 Guo-Wei Wei。结合元素特异性持久同调与机器学习进行蛋白质-配体结合亲和力预测。《生物医学工程数值方法国际期刊》，34(2):e2914，2018。

Gunnar Carlsson。用于数据建模的拓扑方法。《自然综述：物理》，2(12):697–708，2020。

Mathieu Carrière、Marco Cuturi 和 Steve Oudot。Perslay：用于持久图的神经网络层及新的图拓扑特征。载于人工智能与统计国际会议，第2786–2796页，2020年。

Mathieu Carrière、Ulrich Bauer 和 Steve Oudot。使用可微拓扑层优化持久图。国际机器学习会议，2021。

Frédéric Chazal 和 Bertrand Michel。拓扑数据分析导论：面向数据科学家的基础与实践方面。人工智能前沿，4:108，2021。

Frédéric Chazal、Brittany Terese Fasy、Fabrizio Lecci、Alessandro Rinaldo 和 Larry Wasserman。持久景观和轮廓函数的随机收敛。载于第三十届计算几何年度研讨会论文集，第 474–483 页，2014。

Frédéric Chazal、Steve Y. Oudot、Marc Glisse 和 Vin De Silva。持久模的结构与稳定性。施普林格出版社，2016。

Chao Chen 和 Michael Kerber。持久同调的输出敏感算法。载于第二十七届计算几何年度研讨会论文集，第 207–216 页，2011。

Victor Chernozhukov、Denis Chetverikov、Mert Demirer、Esther Duflo、Christian Hansen、Whit- ney Newey 和 James Robins。用于处理效应和结构参数的双重/去偏机器学习。计量经济学杂志，21(1):C1–C68，2018年1月。

Miroslav Dudík、John Langford 和 Lihong Li。双重稳健的策略评估与学习。载于第*28*届国际机器学习国际会议论文集，第1097–1104页，2011年。Kreske Ecker、Xavier de Luna 和 Lina Schelin。函数型结局的因果推断。皇家统计学会会刊*C*辑：应用统计，73(1):221–240，2024年。H. Edelsbrunner 和 J. Harer。计算拓扑学：导论。应用数学。美国数学会，2010年。ISBN 9780821849255。Herbert Edelsbrunner 和 Dmitriy Morozov。26 持久同调。Amirhossein Farzam、Ahmed Aloui、Vahid Tarokh 和 Guillermo Sapiro。用于估计因果效应的拓扑感知鲁棒表示平衡。载于高维学习动态*2025*，2025年。Seyedehsamaneh Gholizadeh 和 Bianca Zadrozny。时间序列与系统分析中的拓扑数据分析简述。*arXiv* 预印本 *arXiv:1809.10745*，2018年。A. Hatcher。代数拓扑。代数拓扑。剑桥大学出版社，2002年。ISBN 9780521795401。Guido W Imbens 和 Donald B Rubin。统计学、社会科学与生物医学中的因果推断。剑桥大学出版社，2015年。Sohail Iqbal、Hafiz Fareed Ahmed、Talha Qaiser、Muhammad Imran Qureshi 和 Nasir Rajpoot。通过 CT 扫描同调分类 COVID-19。《生物与医学中的计算机》，193：110226，2025年。Edward H Kennedy。因果推断中的半参数理论与经验过程。统计因果推断及其在公共卫生研究中的应用，第141–167页，2016年。Edward H Kennedy。基于增量倾向评分干预的非参数因果效应。美国统计学会会刊，114(526):645–656，2019年。Edward H Kennedy。迈向异质性因果效应的最优双重稳健估计。《电子统计学期刊》，17(2):3008–3049，2023年。Edward H Kennedy。半参数双重稳健靶向双重机器学习：综述。精准医学统计方法手册，第207–236页，2024年。Edward H Kennedy、Sivaraman Balakrishnan 和 Max G'sell。用于依从者分类及因果效应推广的精确工具变量。2020年。Edward H Kennedy、Sivaraman Balakrishnan 和 LA Wasserman。半参数反事实密度估计。《生物统计学》，110(4):875–896，2023 年。Kwangho Kim、Jisu Kim 和 Alessandro Rinaldo。通过拓扑数据分析进行时间序列特征化。*arXiv* 预印本 *arXiv:1812.02987*，2018年。Kwangho Kim、Jisu Kim、Manzil Zaheer、Joon Kim、Frédéric Chazal 和 Larry Wasserman。Pllay：基于持久景观的高效拓扑层。神经信息处理系统进展，33:15965–15977，2020年。Violeta Kovacev-Nikolic、Peter Bubenik、Dragan Nikoli´c 和 Giseon Heo。利用持久同调和动力距离分析蛋白质结合。《遗传学与分子生物学中的统计应用》，15(1):19–38，2016年。Dominik Liebl 和 Matthew Reimherr。函数型参数的快速、公平同时置信带。皇家统计学会会刊*B*辑：统计方法论，85(3):842–868，2023年。Whitney K Newey 和 James R Robins。用于半参数估计的交叉拟合与快速余项速率。*arXiv* 预印本 *arXiv:1801.09138*，2018年。

Alessia Pini和Simone Vantini。函数型数据的区间逐点检验。《非参数统计学杂志》，29(2):407–424，2017。Philip T Reiss、Jeff Goldsmith、Han Lin Shang和R Todd Ogden。标量对函数回归方法。《国际统计评论》，85(2):228–249，2017。Ann E Sizemore、Chad Giusti和Danielle S Bassett。整体的重要性：面向网络神经科学家的拓扑数据分析。《网络神经科学》，3(3):656–673，2019年。Eduardo Soares、Plamen Angelov、Sarah Biaso、Michele Higa Froes和Daniel Kanda Abe。SARS-CoV-2 CT扫描数据集：用于SARS-CoV-2识别的大型真实患者CT扫描数据集。*medRxiv*，第2020–04页，2020。Lorenzo Testa、Tobia Boschi、Francesca Chiaromonte、Edward H Kennedy和Matthew Reimherr。双重稳健函数型平均处理效应估计。*arXiv*预印本 *arXiv:2501.06024*，2025。GUDHI 项目。*GUDHI*用户与参考手册。GUDHI编辑委员会，3.4.1版，2021。URL https://gudhi.inria.fr/doc/3.4.1/。Katharine Turner、Sayan Mukherjee和Doug M Boyer。用于形状与曲面建模的持久同调变换。《信息与推断：*IMA*期刊》，3(4):310–344，2014。Aad W Van Der Vaart。半参数统计。载于《概率论与统计学讲座》（圣弗卢尔，*1999*），第331–457页。施普林格，2002。Aad W Van Der Vaart和Jon A Wellner。弱收敛。载于《弱收敛与经验过程》，第16–28页。施普林格，1996年。

附录

标准滤过构造及相关描述符

本附录提供滤流构造的具体定义与示例，以及第2节（滤过的选择）中所引用的相关拓扑摘要。

Vietoris–Rips复形。 设 X 是一个有限点集，并配备距离函数 d（例如，当 $X \subset \mathbb{R}^m$时的欧氏距离）。对于 $r > 0$，尺度为 r 时的 Vietoris–Rips 复形是单纯复形

$$\mathrm{Rips}(r) = \Big\{\sigma \subseteq X : \mathsf{d}(u, u') \le 2r \text{ for all } u, u' \in \sigma\Big\}.$$

若 $r_1 \le r_2$，则 Rips(r_1) $\subseteq$ Rips(r_2)，因此 {Rips(r) : $r \ge 0$} 构成一个自然的滤过。

α-复形。 设 $X \subset \mathbb{R}^m$ 是一个有限点集。对于每个 $u \in X$，令 V_u 表示 u的 Voronoi胞腔，

$$V_u = \big\{x \in \mathbb{R}^m : \|x - u\| \le \|x - u'\| \text{ for all } u' \in X\big\},$$

并令 $B_u(r) = \{x \in \mathbb{R}^m : \|x - u\| \le r\}$ 表示以 u为中心、半径为 r 的闭球。定义截断球 $R_u(r) = B_u(r) \cap V_u$。尺度为 r 时的 α-复形为

$$\mathrm{Alpha}(r) = \Big\{\sigma \subseteq X : \bigcap_{u \in \sigma} R_u(r) \neq \varnothing\Big\}.$$

随着 r 增大，Alpha(r) 呈嵌套关系，从而得到滤过 {Alpha(r) : $r \ge 0$}。

立方复形。立方复形是由轴对齐立方体构成的单纯复形的类似物，非常适用于网格结构数据，例如数字图像。基本区间是形如 $[l, l]$ 或 $[l, l+1]$ 的区间，其中 $l \in \mathbb{Z}$；它们分别称为退化和非退化区间。基本立方体是由基本区间构成的有限乘积 $Q = I_1 \times \cdots \times I_n$，其维数等于非退化因子的数量。立方复形 是由有限个基本立方体组成的集合，并满足：其中任一立方体的每个面也属于 K 中的 K。通过为立方体赋予滤值并取子水平（或超水平）子复形，可以构造滤过立方复形。

持久景观。持久景观（Bubenik等人，2015）将持久图嵌入函数希尔伯特空间。给定一个持久图 $\mathcal{D}$，其持久景观是由 $\{\lambda_k(\cdot; \mathcal{D})\}_{k \in \mathbb{N}}$ 定义的集合：

$$\lambda_k(t; \mathcal{D}) = \operatorname*{kmax}_{p \in \mathcal{D}} \Lambda_p(t), \qquad k \in \mathbb{N},\ t \in \mathbb{T},$$

其中， Λ_p 是方程 1 中的帐篷函数，kmax 表示第 k 大的值（计入重数）， $\mathbb{T} \subset \mathbb{R}$ 是一个固定的紧区间。当 k 固定时，我们可以将 $\lambda(t; \mathcal{D},\ k)$ 记作 $\lambda_k(t;\ \mathcal{D})$，并将 $\lambda(t; \mathcal{D}_d,\ k)$ 记作与第 d 个图对应的景观。若 $\mathcal{D}$ 有 N 个非对角点，则对所有 $k > N$， $\lambda_k(\cdot;\ \mathcal{D}) \equiv 0$。

B 偏差与方差分析

本节探讨当讨厌参数估计 $\hat{\mu}_a(t, x; d)$ 和 $\hat{\pi}_a(x)$ 偏离真实值时，我们的估计量 $\psi_{\mathrm{PI},d}$, $\psi_{\mathrm{IPW},d}$ 和 $\psi_{\mathrm{AIPW},d}$ 的表现。为此，我们仿照 Dudík 等人（2011）的做法，分析这些估计量的偏差和方差。下文省略所有记号中的同调维数 d，因为后续结果均不依赖于 d。令 $\Delta_a(t, x)$ 表示 $\hat{\mu}_a(t, x)$ 相对于 $\mu_a(t, x) = \mathbb{E}(\phi(t;\ \mathcal{D}) \mid X = x, A = a)$ 的加性偏差，并令 $\delta_a(x)$ 表示 $\hat{\pi}_a(x)$ 相对于 $\pi_a(x) = \mathbb{P}(A = a \mid X = x)$ 的乘性偏差，使得

$$\Delta_a(t, x) = \hat{\mu}_a(t, x) - \mu_a(t, x),$$

$$\delta_a(x) = 1 - \frac{\pi_a(x)}{\hat{\pi}_a(x)}.$$

如注释 2 所述，我们假设估计量 $\hat{\mu}_a(t, x)$ 和 $\hat{\pi}_a(x)$ 是通过样本分割从一个独立样本中拟合得到的，因此是固定的。为方便记号，我们用 μ_a 表示 $\mu_a(t, x)$，用 $\hat{\mu}_a$ 表示 $\hat{\mu}_a(t, x)$，用 π_a 表示 $\pi_a(x)$，用 $\hat{\pi}_a$ 表示 $\hat{\pi}_a(x)$，用 $\mathbb{1}_a$ 表示 $\mathbb{1}(A = a)$，用 ϕ 表示$\phi(t; \mathcal{D})$，用 Δ_a 表示 $\Delta_a(t, x)$，并用 δ_a 表示 $\delta_a(x)$。

B.1 偏差分析

我们首先分析 $\widehat{\psi}_{\mathrm{AIPW}}$ 的偏差。令 $\varphi_a := \mu_a + \frac{\mathbb{1}_a}{\pi_a}(\phi - \mu_a)$，并令 $\widehat{\varphi}_a$ 为使用 $\hat{}\mu_a$ 和 $\widehat{\pi}_a$ 得到的估计量。那么，$\widehat{\varphi}_a$ 的期望为

$$\begin{aligned}
\mathbb{E}(\widehat{\varphi}_a) &= \mathbb{E}\left\{\hat{\mu}_a + \frac{\mathbb{1}_a}{\hat{\pi}_a}(\phi - \hat{\mu}_a)\right\} \\
&= \mathbb{E}\left\{\mu_a + \Delta_a + \frac{\mathbb{1}_a}{\hat{\pi}_a}(\phi - \mu_a - \Delta_a)\right\} \\
&= \mathbb{E}(\mu_a) + \mathbb{E}_X\left[\mathbb{E}\left\{\Delta_a + \frac{\mathbb{1}_a}{\hat{\pi}_a}(\phi - \mu_a - \Delta_a)\middle| X\right\}\right] \\
&= \mathbb{E}(\phi^a) + \mathbb{E}_X\left\{\Delta_a\left(1 - \frac{\pi_a}{\hat{\pi}_a}\right) + \frac{\pi_a}{\hat{\pi}_a}(\mu_a - \mu_a)\right\} \\
&= \mathbb{E}(\phi^a) + \mathbb{E}(\Delta_a\delta_a),
\end{aligned}$$

其中，第四个等式由全期望定律，以及第3节中的（C1）一致性和（C2）无混杂性假设共同推出。根据上述表述，我们可以推导出$\widehat{}$的 ψ_{AIPW} 偏差如下：

$$\begin{aligned}
\mathbb{E}\left(\widehat{\psi}_{\mathrm{AIPW}}\right) - \psi &= \mathbb{E}(\widehat{\varphi}_1 - \widehat{\varphi}_0) - \psi \\
&= \mathbb{E}(\phi^1 - \phi^0) - \psi + \mathbb{E}(\Delta_1\delta_1 - \Delta_0\delta_0) \\
&= \mathbb{E}(\Delta_1\delta_1 - \Delta_0\delta_0).
\end{aligned}$$

$\widehat{\quad}$ $\widehat{\quad}$ ψ_{PI} 和 ψ_{IPW} 的偏差自然可由以下观察得到：PI 估计量和逆概率加权估计量分别是增广逆概率加权估计量的特殊情形，即$-\frac{\mathbb{1}_a}{\hat{\pi}_a}(\phi - \hat{}\mu_a) = 0$ 和 $\hat{}\mu_a = 0$。

$$\begin{aligned}
\mathbb{E}\left(\widehat{\psi}_{\mathrm{PI}}\right) - \psi &= \mathbb{E}(\Delta_1 - \Delta_0) \\
\mathbb{E}\left(\widehat{\psi}_{\mathrm{IPW}}\right) - \psi &= -\mathbb{E}(\delta_1\mu_1 - \delta_0\mu_0).
\end{aligned}$$

一般而言，就偏差而言，没有任何一个估计量能够一致地优于其他估计量。然而，如果任一 $\hat{}\mu_a \approx \mu_a$ 或 $\hat{}\pi_a \approx \pi_a$成立，则增广逆概率加权估计量的偏差将接近 0；而 PI 估计量通常需要 $\hat{}\mu_a \approx \mu_a$，逆概率加权估计量通常需要 $\hat{}\pi_a \approx \pi_a$。因此，增广逆概率加权估计量能够有效整合两种信息来源，从而实现更好的估计。

B.2 方差分析

$\hat{}$ $\hat{}$ $\hat{}$ 我们现在分析估计量的方差。为方便起见，对任意 $s, t \in \mathbb{T}$，我们令 $\mathrm{Var}(\psi) := \mathrm{Cov}(\psi(s), \psi(t))$，且后续结果对 s 与 t 的任意选取均成立。为简化记号，我们定义 $\epsilon_a := \frac{\mathbb{1}_a}{\hat{\pi}_a}(\phi - \mu_a)$。注意，由全期望定律，$\mathbb{E}(\epsilon_a \mid X) = 0$。

增广逆概率加权。 鉴于 $\mathrm{Var}(\hat{}\psi_{\mathrm{AIPW}}) = \frac{1}{n}\mathrm{Var}(\hat{}\varphi_1 - \hat{}\varphi_0)$，只需分析 $\mathrm{Var}(\hat{}\varphi_1 - \hat{}\varphi_0)$。由于 $\mathrm{Var}(\hat{}\varphi_1 - \hat{}\varphi_0) = \mathbb{E}\{(\hat{}\varphi_1 - \hat{}\varphi_0)^2\} - \mathbb{E}(\hat{}\varphi_1 - \hat{}\varphi_0)^2$ 且 $\mathbb{E}(\hat{}\varphi_1 - \hat{}\varphi_0)$ 已在上一节推导，因此分析 $\mathbb{E}\{(\hat{}\varphi_1 - \hat{}\varphi_0)^2\}$ 即可得到所需结果。

$$\begin{aligned}
\mathbb{E}\{(\hat{\varphi}_1 - \hat{\varphi}_0)^2\} &= \mathbb{E}\left[\left\{\hat{\mu}_1 + \frac{\mathbb{1}_1}{\hat{\pi}_1}(\phi - \hat{\mu}_1) - \hat{\mu}_0 - \frac{\mathbb{1}_0}{\hat{\pi}_0}(\phi - \hat{\mu}_0)\right\}^2\right] \\
&= \mathbb{E}\left[\left\{(\mu_1 - \mu_0) + (\epsilon_1 - \epsilon_0) + \Delta_1\left(1 - \frac{\mathbb{1}_1}{\hat{\pi}_1}\right) - \Delta_0\left(1 - \frac{\mathbb{1}_0}{\hat{\pi}_0}\right)\right\}^2\right] \\
&= \mathbb{E}\left\{(\mu_1 - \mu_0)^2\right\} + \mathbb{E}\left\{(\epsilon_1 - \epsilon_0)^2\right\} + \mathbb{E}\left[\left\{\Delta_1\left(1 - \frac{\mathbb{1}_1}{\hat{\pi}_1}\right) - \Delta_0\left(1 - \frac{\mathbb{1}_0}{\hat{\pi}_0}\right)\right\}^2\right] \\
&\quad + 2\mathbb{E}\left[(\mu_1 - \mu_0)\left\{\Delta_1\left(1 - \frac{\mathbb{1}_1}{\hat{\pi}_1}\right) - \Delta_0\left(1 - \frac{\mathbb{1}_0}{\hat{\pi}_0}\right)\right\}\right],
\end{aligned}$$

其中， $\epsilon_1 - \epsilon_0$ 的交叉乘积的期望为 0，这是迭代期望以及$\mathbb{E}(\epsilon_a \mid X) = 0$的结果。接下来，我们进一步展开第三项和最后一项。第三项可展开为

$$
\begin{aligned}
&\mathbb{E}\left[\left\{\Delta_1\left(1-\frac{\mathbb{1}_1}{\hat{\pi}_1}\right)-\Delta_0\left(1-\frac{\mathbb{1}_0}{\hat{\pi}_0}\right)\right\}^2\right] \\
&=\mathbb{E}\left\{\Delta_1^2\left(1-\frac{\mathbb{1}_1}{\hat{\pi}_1}\right)^2+\Delta_0^2\left(1-\frac{\mathbb{1}_0}{\hat{\pi}_0}\right)^2-2\Delta_1\Delta_0\left(1-\frac{\mathbb{1}_1}{\hat{\pi}_1}\right)\left(1-\frac{\mathbb{1}_0}{\hat{\pi}_0}\right)\right\} \\
&=\mathbb{E}\left\{\Delta_1^2\left(1-2\frac{\mathbb{1}_1}{\hat{\pi}_1}+\frac{\mathbb{1}_1}{\hat{\pi}_1^2}\right)+\Delta_0^2\left(1-2\frac{\mathbb{1}_0}{\hat{\pi}_0}+\frac{\mathbb{1}_0}{\hat{\pi}_0^2}\right)-2\Delta_1\Delta_0\left(1-\frac{\mathbb{1}_1}{\hat{\pi}_1}-\frac{\mathbb{1}_0}{\hat{\pi}_0}\right)\right\} \\
&=\mathbb{E}_X\left\{\Delta_1^2\left(1-2\frac{\pi_1}{\hat{\pi}_1}+\frac{\pi_1}{\hat{\pi}_1^2}\right)+\Delta_0^2\left(1-2\frac{\pi_0}{\hat{\pi}_0}+\frac{\pi_0}{\hat{\pi}_0^2}\right)-2\Delta_1\Delta_0\left(1-\frac{\pi_1}{\hat{\pi}_1}-\frac{\pi_0}{\hat{\pi}_0}\right)\right\} \\
&=\mathbb{E}_X\left\{\Delta_1^2\delta_1^2+\Delta_1^2\left(\frac{\pi_1(1-\pi_1)}{\hat{\pi}_1^2}\right)+\Delta_0^2\delta_0^2+\Delta_0^2\left(\frac{\pi_0(1-\pi_0)}{\hat{\pi}_0^2}\right)-2\Delta_1\Delta_0\delta_1\delta_0+2\Delta_1\Delta_0\frac{\pi_1\pi_0}{\hat{\pi}_1\hat{\pi}_0}\right\} \\
&=\mathbb{E}_X\left\{(\Delta_1\delta_1-\Delta_0\delta_0)^2+\Delta_1^2\left(\frac{\pi_1\pi_0}{\hat{\pi}_1^2}\right)+\Delta_0^2\left(\frac{\pi_1\pi_0}{\hat{\pi}_0^2}\right)+2\Delta_1\Delta_0\frac{\pi_1\pi_0}{\hat{\pi}_1\hat{\pi}_0}\right\} \\
&=\mathbb{E}_X\left\{(\Delta_1\delta_1-\Delta_0\delta_0)^2+\Delta_1^2(1-\delta_1)^2\left(\frac{\pi_0}{\pi_1}\right)+\Delta_0^2(1-\delta_0)^2\left(\frac{\pi_1}{\pi_0}\right)+2\Delta_1\Delta_0(1-\delta_1)(1-\delta_0)\right\} \\
&=\mathbb{E}_X\left\{(\Delta_1\delta_1-\Delta_0\delta_0)^2\right\}+\mathbb{E}_X\left[\left\{\Delta_1(1-\delta_1)\sqrt{\pi_0/\pi_1}+\Delta_0(1-\delta_0)\sqrt{\pi_1/\pi_0}\right\}^2\right],
\end{aligned}
$$

其中，第三个等式由迭代期望得到。最后一项可展开为

$$
\begin{aligned}
\mathbb{E}\left[(\mu_1-\mu_0)\left\{\Delta_1\left(1-\frac{\mathbb{1}_1}{\hat{\pi}_1}\right)-\Delta_0\left(1-\frac{\mathbb{1}_0}{\hat{\pi}_0}\right)\right\}\right] &= \mathbb{E}_X\left[(\mu_1-\mu_0)\left\{\Delta_1\left(1-\frac{\pi_1}{\hat{\pi}_1}\right)-\Delta_0\left(1-\frac{\pi_0}{\hat{\pi}_0}\right)\right\}\right] \\
&= \mathbb{E}_X\{(\mu_1-\mu_0)(\Delta_1\delta_1-\Delta_0\delta_0)\},
\end{aligned}
$$

其中，第一个等式由迭代期望得到。结合上述结果，我们有

$$
\begin{aligned}
\mathbb{E}\{(\hat{\varphi}_1-\hat{\varphi}_0)^2\} &= \mathbb{E}\left\{(\mu_1-\mu_0)^2\right\}+2\mathbb{E}_X\{(\mu_1-\mu_0)(\Delta_1\delta_1-\Delta_0\delta_0)\}+\mathbb{E}_X\left\{(\Delta_1\delta_1-\Delta_0\delta_0)^2\right\} \\
&\quad+\mathbb{E}\left\{(\epsilon_1-\epsilon_0)^2\right\}+\mathbb{E}_X\left[\left\{\Delta_1(1-\delta_1)\sqrt{\pi_0/\pi_1}+\Delta_0(1-\delta_0)\sqrt{\pi_1/\pi_0}\right\}^2\right] \\
&= \mathbb{E}\left\{(\mu_1+\Delta_1\delta_1-\mu_0-\Delta_0\delta_0)^2\right\}+\mathbb{E}\left\{(\epsilon_1-\epsilon_0)^2\right\} \\
&\quad+\mathbb{E}_X\left[\left\{\Delta_1(1-\delta_1)\sqrt{\pi_0/\pi_1}+\Delta_0(1-\delta_0)\sqrt{\pi_1/\pi_0}\right\}^2\right].
\end{aligned}
$$

在推导出 $\mathbb{E}\{(\hat{\varphi}_1-\hat{\varphi}_0)^2\}$ 之后，我们现在可以计算增广逆概率加权估计量的方差。根据上一节的偏差分析，我们知道 $\mathbb{E}(\hat{\varphi}_1-\hat{\varphi}_0) = \mathbb{E}(\mu_1+\Delta_1\delta_1-\mu_0-\Delta_0\delta_0)$。因此，增广逆概率加权估计量的方差为

$$
\begin{aligned}
\mathrm{Var}(\hat{\psi}_{\mathrm{AIPW}}) &= \frac{1}{n}\mathrm{Var}(\hat{\varphi}_1-\hat{\varphi}_0) \\
&= \frac{1}{n}\Big(\mathbb{E}\{(\hat{\varphi}_1-\hat{\varphi}_0)^2\}-\mathbb{E}(\hat{\varphi}_1-\hat{\varphi}_0)^2\Big) \\
&= \frac{1}{n}\Bigg(\mathrm{Var}\left(\mu_1+\Delta_1\delta_1-\mu_0-\Delta_0\delta_0\right)+\mathbb{E}\left\{(\epsilon_1-\epsilon_0)^2\right\} \\
&\quad+\mathbb{E}_X\left[\left\{\Delta_1(1-\delta_1)\sqrt{\pi_0/\pi_1}+\Delta_0(1-\delta_0)\sqrt{\pi_1/\pi_0}\right\}^2\right]\Bigg).
\end{aligned}
$$

ˆ逆概率加权。为了推导 ψ_{IPW} 的方差，我们采用类似的方法，首先分析二阶矩。

$$\begin{aligned}
\mathbb{E}\left\{\left(\frac{\mathbb{1}_1}{\hat{\pi}_1}\phi-\frac{\mathbb{1}_0}{\hat{\pi}_0}\phi\right)^2\right\} &= \mathbb{E}\left\{\left(\epsilon_1-\epsilon_0+\frac{\mathbb{1}_1}{\hat{\pi}_1}\mu_1-\frac{\mathbb{1}_0}{\hat{\pi}_0}\mu_0\right)^2\right\} \\
&= \mathbb{E}\left[\left\{(\epsilon_1-\epsilon_0)+(\mu_1-\mu_0)-\mu_1\left(1-\frac{\mathbb{1}_1}{\hat{\pi}_1}\right)+\mu_0\left(1-\frac{\mathbb{1}_0}{\hat{\pi}_0}\right)\right\}^2\right] \\
&= \mathbb{E}\left\{(\epsilon_1-\epsilon_0)^2\right\}+\mathbb{E}\left\{(\mu_1-\mu_0)^2\right\}+\mathbb{E}\left[\left\{\mu_1\left(1-\frac{\mathbb{1}_1}{\hat{\pi}_1}\right)-\mu_0\left(1-\frac{\mathbb{1}_0}{\hat{\pi}_0}\right)\right\}^2\right] \\
&\quad -2\mathbb{E}\left[(\mu_1-\mu_0)\left\{\mu_1\left(1-\frac{\mathbb{1}_1}{\hat{\pi}_1}\right)-\mu_0\left(1-\frac{\mathbb{1}_0}{\hat{\pi}_0}\right)\right\}\right],
\end{aligned}$$

根据迭代期望，$\epsilon_1-\epsilon_0$ 的交叉乘积的期望为 0，且$\mathbb{E}(\epsilon_a \mid X)=0$。注意，第三项和最后一项的结构与推导增广逆概率加权估计量时出现的相应项相同。因此，可以直接应用前述结果，从而得到

$$\mathbb{E}\left[\left\{\mu_1\left(1-\frac{\mathbb{1}_1}{\hat{\pi}_1}\right)-\mu_0\left(1-\frac{\mathbb{1}_0}{\hat{\pi}_0}\right)\right\}^2\right]=\mathbb{E}_X\left\{(\mu_1\delta_1-\mu_0\delta_0)^2\right\}+\mathbb{E}_X\left[\left\{\mu_1(1-\delta_1)\sqrt{\pi_0/\pi_1}+\mu_0(1-\delta_0)\sqrt{\pi_1/\pi_0}\right\}^2\right],$$

and

$$\mathbb{E}\left[(\mu_1-\mu_0)\left\{\mu_1\left(1-\frac{\mathbb{1}_1}{\hat{\pi}_1}\right)-\mu_0\left(1-\frac{\mathbb{1}_0}{\hat{\pi}_0}\right)\right\}\right]=\mathbb{E}_X\{(\mu_1-\mu_0)(\mu_1\delta_1-\mu_0\delta_0)\}.$$

结合上述结果，我们得到

$$\begin{aligned}
\mathbb{E}\left\{\left(\frac{\mathbb{1}_1}{\hat{\pi}_1}\phi-\frac{\mathbb{1}_0}{\hat{\pi}_0}\phi\right)^2\right\} &= \mathbb{E}\left\{(\mu_1-\mu_0)^2\right\}-2\mathbb{E}_X\{(\mu_1-\mu_0)(\mu_1\delta_1-\mu_0\delta_0)\}+\mathbb{E}_X\left\{(\mu_1\delta_1-\mu_0\delta_0)^2\right\} \\
&\quad +\mathbb{E}\left\{(\epsilon_1-\epsilon_0)^2\right\}+\mathbb{E}_X\left[\left\{\mu_1(1-\delta_1)\sqrt{\pi_0/\pi_1}+\mu_0(1-\delta_0)\sqrt{\pi_1/\pi_0}\right\}^2\right] \\
&= \mathbb{E}\left[\left\{(1-\delta_1)\mu_1-(1-\delta_0)\mu_0\right\}^2\right]+\mathbb{E}\left\{(\epsilon_1-\epsilon_0)^2\right\} \\
&\quad +\mathbb{E}_X\left[\left\{\mu_1(1-\delta_1)\sqrt{\pi_0/\pi_1}+\mu_0(1-\delta_0)\sqrt{\pi_1/\pi_0}\right\}^2\right].
\end{aligned}$$

根据上一节的偏差分析，我们知道 $\mathbb{E}\left(\frac{\mathbb{1}_1}{\hat{\pi}_1}\phi-\frac{\mathbb{1}_0}{\hat{\pi}_0}\phi\right)=\mathbb{E}\{(1-\delta_1)\mu_1-(1-\delta_0)\mu_0\}$。因此，逆概率加权估计量的方差为

$$\begin{aligned}
\mathrm{Var}(\hat{\psi}_{\mathrm{IPW}}) &= \frac{1}{n}\mathrm{Var}\left(\frac{\mathbb{1}_1}{\hat{\pi}_1}\phi-\frac{\mathbb{1}_0}{\hat{\pi}_0}\phi\right) \\
&= \frac{1}{n}\left(\mathbb{E}\left\{\left(\frac{\mathbb{1}_1}{\hat{\pi}_1}\phi-\frac{\mathbb{1}_0}{\hat{\pi}_0}\phi\right)^2\right\}-\mathbb{E}\left(\frac{\mathbb{1}_1}{\hat{\pi}_1}\phi-\frac{\mathbb{1}_0}{\hat{\pi}_0}\phi\right)^2\right) \\
&= \frac{1}{n}\Bigg(\mathrm{Var}\left\{(1-\delta_1)\mu_1-(1-\delta_0)\mu_0\right\}+\mathbb{E}\left\{(\epsilon_1-\epsilon_0)^2\right\} \\
&\quad +\mathbb{E}_X\left[\left\{\mu_1(1-\delta_1)\sqrt{\pi_0/\pi_1}+\mu_0(1-\delta_0)\sqrt{\pi_1/\pi_0}\right\}^2\right]\Bigg).
\end{aligned}$$

ˆ ˆ 我们注意到，$\mathrm{Var}(\psi_{\mathrm{AIPW}})$ 和 $\mathrm{Var}(\psi_{\mathrm{IPW}})$ 的结构相似。当 $\hat{\pi}\approx\pi$ 时，对于两种估计量，第一项通常量级相近；但如果同时满足 $\hat{\mu}\approx\mu$，AIPW 的第一项可能更小。第二项完全相同。第三项的唯一区别是，AIPW 涉及 Δ，而 IPW 则将其替换为 μ。一般而言，Δ 的量级会小于 μ，因此 IPW 比 AIPW 更可能表现出较大的方差。

ˆ插件估计。 ψ_{PI} 的方差很容易求得：

$$\begin{aligned}\mathrm{Var}(\hat{\psi}_{\mathrm{PI}}) &= \frac{1}{n}\mathrm{Var}(\hat{\mu}_1 - \hat{\mu}_0) \\ &= \frac{1}{n}\mathrm{Var}(\mu_1 + \Delta_1 - \mu_0 - \Delta_0).\end{aligned}$$

与增广逆概率加权和逆概率加权估计量不同，PI 的方差只包含一项，对应于增广逆概率加权和逆概率加权方差分解中的第一部分。因此，PI 估计量的方差通常小于增广逆概率加权和逆概率加权估计量。不过，正如偏差分析中所讨论的，如果回归模型设定错误，PI 估计量容易产生较大的偏差。

C 证明

记号。对于任意固定向量 x，我们用 $\|x\|_q$ 表示 L_q 范数。对于给定函数 f，我们用记号 $\|f\|_{\mathbb{P},q} = [\mathbb{P}(|f|^q)]^{1/q} = \left[\int |f(z)|^q d\mathbb{P}(z)\right]^{1/q}$表示 f的 $L_q(\mathbb{P})$-范数。另外，我们用 $\mathbb{P}$表示给定样本算子 $\hat{}$ f 后的条件期望，如 $\mathbb{P}(\hat{}f) = \int \hat{f}(z) d\mathbb{P}(z)$ 所示。注意，$\hat{}\hat{}\hat{}$ 只有当 f依赖于样本时，$\mathbb{P}(f)$ 才是随机的，此时 $\mathbb{P}(f) \neq \mathbb{E}(f)$。否则，$\mathbb{P}\hat{}$ 和 $\mathbb{E}$ 可以互换使用。例如，若 f 是在独立的（训练）样本 $\mathbf{D}^n = (Z_1, ..., Z_n)$ 上构造的，那么对于一个新的观测值 $Z \sim \mathbb{P}$，有 $\mathbb{P}\left\{\hat{f}(Z)\right\} = \mathbb{E}\left\{\hat{f}(Z) \mid \mathbf{D}^n\right\}$。我们用 $\mathbb{P}_n$表示经验测度，如 $\mathbb{P}_n(f) = \mathbb{P}_n\{f(Z)\} = \frac{1}{n}\sum_{i=1}^n f(Z_i)$ 所示。最后，我们用简写 $a_n \lesssim b_n$ 表示某个普适常数 $\mathrm{c} > 0$下的 $a_n \leq \mathrm{c} b_n$。

C.1 引理 2.1 的证明

证明。 为方便记号，对于任意固定的持久图$\mathcal{D}$和幂参数 r，记 $\phi(t) := \phi(t; \mathcal{D}, r)$。固定 $p = (a, b) \in \mathcal{D}$。帐篷函数 Λ_p在 $[a, (a+b)/2]$上是斜率为 $+1$ 的线性函数，在 $[(a+b)/2, b]$ 上是斜率为 -1的线性函数。因此，对所有 $s, t \in \mathbb{R}$，

$$|\Lambda_p(s) - \Lambda_p(t)| \leq |s - t|.$$

也就是说，每个 Λ_p 都是 1-利普希茨的。

将加权轮廓写为 $\phi(t) = \sum_{j=1}^N \alpha_j \Lambda_{p_j}(t)$，其权重为 $\alpha_j = w_j / \sum_k w_k \in (0, 1)$，且满足 $\sum_j \alpha_j = 1$。于是可得

$$|\phi(s) - \phi(t)| = \left|\sum_j \alpha_j \big(\Lambda_{p_j}(s) - \Lambda_{p_j}(t)\big)\right| \leq \sum_j \alpha_j |s - t| = |s - t|,$$

这表明剪影函数也是常数为 $L = 1$的利普希茨函数。

即使持久图 $\mathcal{D}$ 是随机的，该不等式仍逐样本成立；因此，取期望可得

$$\mathbb{E}\Big[\sup_{|s-t| \leq \delta} |\phi(s) - \phi(t)|\Big] \leq \delta.$$

□

C.2 定理 5.1 的证明

证明。 由于 π 已知，我们有

$$\widehat{\psi}_{\mathrm{IPW},d}(t) = \mathbb{P}_n\{\xi_d(Z, t)\}, \qquad \psi_d(t) = \mathbb{P}\{\xi_d(Z, t)\}, \qquad t \in \mathbb{T}.$$

因此，

$$\sqrt{n}\big\{\widehat{\psi}_{\mathrm{IPW},d}(t) - \psi_d(t)\big\} = \sqrt{n}(\mathbb{P}_n - \mathbb{P})\{\xi_d(Z, t)\} = \mathbb{G}_n\, \xi_d(\cdot, t),$$

其中，$\mathbb{G}_n := \sqrt{n}(\mathbb{P}_n - \mathbb{P})$ 表示经验过程。

根据给定条件，指标函数类 $\{\xi_d(\cdot,t):\ t\in\mathbb{T}\}$ 是 $\mathbb{P}$-Donsker（或满足 $\ell^\infty(\mathbb{T})$ 中函数型中心极限定理所需的任何其他标准充分条件），函数型中心极限定理给出

$$\big(\mathbb{G}_n\,\xi_d(\cdot,t)\big)_{t\in\mathbb{T}} \rightsquigarrow \big(\mathbb{G}_d(t)\big)_{t\in\mathbb{T}} \quad \text{in} \quad \ell^\infty(\mathbb{T}),$$

其中，$\mathbb{G}_d$ 是一个以 $\mathbb{T}$为索引的紧均值为零的高斯过程。

另一方面，对于任意 $s,\ t\in\mathbb{T}$，我们有

$$\operatorname{cov}\{\mathbb{G}_d(s),\mathbb{G}_d(t)\}=\operatorname{cov}\big(\xi_d(Z,s),\xi_d(Z,t)\big),$$

whic这可由 Donsk
er 定理以及 $\sup_{t\in\mathbb{T}}\operatorname{var}\{\xi_d(Z,\ t)\}<\infty$ 推得（Van Der Vaart & Wellner，1996，定理 2.5。 [2)].
将上述展示与经验过程表示相结合，即得所需结果：

$$\sqrt{n}\big\{\widehat{\psi}_{\mathrm{IPW},d}(t)-\psi_d(t)\big\}\rightsquigarrow\mathbb{G}_d(t)\quad\text{in}\quad\ell^\infty(\mathbb{T}).$$

□

备注3. 将函数型中心极限定理应用于由 $\{\xi_d(\cdot,t):\ t\in\mathbb{T}\}$ 索引的经验过程的一个充分条件是，该类为 $\mathbb{P}$-*Donsker*，这可由在 $L_2(\mathbb{P})$ 包络及适当可测性条件下的标准括号熵界推出。在我们的设定中，这些条件是合理的，因为 $t\mapsto\phi_{i,d}(t)$ 是利普希茨的，从而可转化为对 $t\mapsto\xi_d(Z,t)$ 的利普希茨对照，并因此得到较小的括号熵，如下一引理所形式化。

引理 C.1 (针对 $\{\xi_d(\cdot,t)\}$的括号熵充分条件)。 假设 $\mathbb{T}\subset\mathbb{R}$ 为紧致集，其直径 $\operatorname{diam}(\mathbb{T})<\infty$。则对所有 $s,t\in\mathbb{T}$,

$$|\xi_d(Z,s)-\xi_d(Z,t)|\le\Big(\frac{1}{\kappa}+\frac{1}{\kappa}\Big)\,|s-t|=\frac{2}{\kappa}|s-t|\quad a.s.$$

此外，令 $F(Z)\coloneqq\sup_{t\in\mathbb{T}}|\xi_d(Z,t)|$，我们有 $F\in L_2(\mathbb{P})$，只要 $\sup_{t\in\mathbb{T}}\operatorname{var}\{\xi_d(Z,t)\}<\infty$ 且 $\sup_{t\in\mathbb{T}}|\mathbb{P}\,\xi_d(Z,t)|<\infty$。在此情形下，括号数满足

$$N_{[]}(\varepsilon\|F\|_{\mathbb{P},2},\,\{\xi_d(\cdot,t):t\in\mathbb{T}\},\,L_2(\mathbb{P}))\ \le\ C\,\varepsilon^{-1},\qquad 0<\varepsilon\le 1,$$

其中常数 C 仅取决于$\operatorname{diam}(\mathbb{T})$ 和 κ。因此，

$$\int_0^1\sqrt{\log N_{[]}(\varepsilon\|F\|_{\mathbb{P},2},\{\xi_d(\cdot,t)\},L_2(\mathbb{P}))}\,d\varepsilon\ <\ \infty,$$

因此，根据标准的括号熵判据，$\{\xi_d(\cdot,t):\ t\in\mathbb{T}\}$ 是 $\mathbb{P}$-*Donsker*。

证明。 固定 $s,t\in\mathbb{T}$。根据定义，

$$\xi_d(Z,t)=\left\{\frac{A}{\pi(X)}-\frac{1-A}{1-\pi(X)}\right\}\phi_{i,d}(t).$$

在正值性条件 $\kappa\le\pi(X)\le1-\kappa$ 几乎处处成立时，当前置因子满足 $A=1$ 时，其绝对值不超过$1/\kappa$；当 $A=0$时，其绝对值不超过 $1/\kappa$，因此一致地不超过 $2/\kappa$。根据引理 2.1，我们得到

$$|\xi_d(Z,s)-\xi_d(Z,t)|\le\left|\frac{A}{\pi(X)}-\frac{1-A}{1-\pi(X)}\right|\cdot|\phi_{i,d}(s)-\phi_{i,d}(t)|\le\frac{2}{\kappa}|s-t|.$$

现在用网格 $\{t_j\}_{j=1}^m$ 覆盖 $\mathbb{T}$，其网格尺寸至多为 $\delta=\varepsilon\|F\|_{\mathbb{P},2}\kappa/2$，使得 $m\le1+\operatorname{diam}(\mathbb{T})/\delta$。对每个 $t\in\mathbb{T}$，选取 $j(t)$，使得 $|t-t_{j(t)}|\le\delta$，并定义括号

$$\ell_t(Z)=\xi_d(Z,t_{j(t)})-\frac{2}{\kappa}\delta,\qquad u_t(Z)=\xi_d(Z,t_{j(t)})+\frac{2}{\kappa}\delta.$$

于是 $\ell_t(Z)\ \le\ \xi_d(Z,t)\ \le\ u_t(Z)$几乎必然成立，且$\|u_t-\ell_t\|_{\mathbb{P},2}\ \le(4/\kappa)\delta\ =\ 2\varepsilon\|F\|_{\mathbb{P},2}$。因此$N_{[]}(\varepsilon\|F\|_{\mathbb{P},2},\cdot,L_2(\mathbb{P}))\ \le\ m\ \lesssim\ \varepsilon^{-1}$，从而证明了熵界，进而证明了有限括号积分。

□

C.3 在假设 (A5′) 下定理 5.2 的证明

回顾可知， φ 是方程3)中目标泛函的非中心化有效影响函数，并满足冯·米塞斯展开：

$$\psi(t;\widehat{\mathbb{P}}) - \psi(t;\mathbb{P}) = \int \varphi(t,z;\eta)\, d(\widehat{\mathbb{P}} - \mathbb{P})(z) + R_2(\widehat{\mathbb{P}}, \mathbb{P}), \tag{10}$$

其中二阶余项规定为

$$\begin{aligned} R_2(\widehat{\mathbb{P}}, \mathbb{P}) = &\int \left\{ \frac{1}{\pi(X)} - \frac{1}{\hat{\pi}(X)} \right\} \{\mu_1(t,X;d) - \hat{\mu}_1(t,X;d)\}\, \pi(X)\, d\mathbb{P} \\ &- \int \left\{ \frac{1}{1-\pi(X)} - \frac{1}{1-\hat{\pi}(X)} \right\} \{\mu_0(t,X;d) - \hat{\mu}_0(t,X;d)\}\, \{1-\pi(X)\}\, d\mathbb{P}, \forall t \in \mathbb{T}. \end{aligned} \tag{11}$$

首先，我们证明如下引理所述的有限维因果效应投影的渐近正态性。该证明遵循 van der Vaart（2002）中定理 5.31 的论证；Kennedy（2019年）也采用了与之类似的技术。

引理 C.2 （离散时间估计量的渐近正态性）。 设 $k \in \mathbb{N}$ 与 t_1，……， $t_k \in \mathcal{T}$ 固定。在整个证明过程中，我们记$\psi_d = \psi_{\mathrm{AIPW},\, d}$。定义

$$\widehat{\boldsymbol{\psi}}_{t_1,\ldots,t_k} = \big(\widehat{\psi}_d(t_1), \ldots, \widehat{\psi}_d(t_k)\big)^\top, \qquad \boldsymbol{\psi}_{t_1,\ldots,t_k} = \big(\psi_d(t_1), \ldots, \psi_d(t_k)\big)^\top.$$

在假设 *(A1)-(A5)* 下，我们有

$$\sqrt{n}\left(\widehat{\boldsymbol{\psi}}_{t_1,\ldots,t_k} - \boldsymbol{\psi}_{t_1,\ldots,t_k}\right) \xrightarrow{d} N(0, \Sigma_{t_1,\ldots,t_k}),$$

其中协方差矩阵为

$$\Sigma_{t_1,\ldots,t_k} = \mathbb{E}\left[\boldsymbol{\varphi}_{t_1,\ldots,t_k}(\mathcal{D})\boldsymbol{\varphi}_{t_1,\ldots,t_k}(\mathcal{D})^\top\right],$$

其中我们记为 $\boldsymbol{\varphi}_{t_1,\ldots,t_k}(\mathcal{D}) = (\varphi(t_1;\, \mathcal{D}) \ldots, \varphi(t_k;\, \mathcal{D}))^\top$。

证明。 为简化记号，令 $\boldsymbol{\varphi} = \boldsymbol{\varphi}_{t_1,\ldots,t_k}(\mathcal{D})$ 和 $\widehat{\boldsymbol{\varphi}} = \widehat{\boldsymbol{\varphi}}_{t_1,\ldots,t_k}(\mathcal{D}) = (\widehat{\varphi}(t_1;\, \mathcal{D}) \ldots, \widehat{\varphi}(t_k;\, \mathcal{D}))^\top$。回顾一下， $\widehat{\varphi}(t;\, \mathcal{D})$ 定义为

$$\widehat{\varphi}(t;\mathcal{D}) = \hat{\mu}_1(t,X;d) - \hat{\mu}_0(t,X;d) + \left\{ \frac{A}{\hat{\pi}(X)} - \frac{1-A}{1-\hat{\pi}(X)} \right\} \{\phi(t;\mathcal{D}_d) - \hat{\mu}_A(t,X;d)\},$$

其中所有干扰参数估计量均基于单独的独立样本构造（假设（A2））。接下来，考虑如下分解：

$$\sqrt{n}\left(\widehat{\boldsymbol{\psi}}_{t_1,\ldots,t_k} - \boldsymbol{\psi}_{t_1,\ldots,t_k}\right) = \sqrt{n}(\mathbb{P}_n - \mathbb{P})\boldsymbol{\varphi} + \sqrt{n}(\mathbb{P}_n - \mathbb{P})(\widehat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}) + \mathbb{P}(\widehat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}). \tag{12}$$

根据多元中心极限定理，公式 12 中的第一项收敛于均值为 0、协方差矩阵为 $\Sigma_{t_1,\ldots,t_k}$ 的多元正态分布。公式 12 中的第三项仅为 $\sqrt{n}R_2$，其中 R_2 是公式 11 中指定的余项。在假设（A4）的速率条件下，可得 $\sqrt{n}R_2 = o_{\mathbb{P}}(1)$。因此，只需证明第二个经验过程项可以忽略不计，即其阶为 $o_{\mathbb{P}}(1)$。根据一致性条件（A3）和三角不等式，

$$\|\widehat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}\|_{\mathbb{P},2} = O\left(\sum_{j=1}^{k} \|\widehat{\varphi}(t_j) - \varphi(t_j)\|_{\mathbb{P},2}\right) = o_{\mathbb{P}}(1).$$

将多维切比雪夫不等式应用于 Kennedy 等人（2020）中引理 2 的证明，可以立即得到

$$(\mathbb{P}_n - \mathbb{P})(\widehat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}) = O_{\mathbb{P}}\left(\frac{\|\widehat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}\|_{\mathbb{P},2}}{\sqrt{n}}\right) = o_{\mathbb{P}}(n^{-1/2}).$$

将所有部分代入公式 12，即可由斯卢茨基定理得出结果。 □

现在我们可以证明定理 5.2。在此，我们采用假设（A5′），并遵循 Testa 等人（2025）的证明策略。随后，我们将在 C.4 节证明，在假设（A5）的较弱条件下，同样的结果仍然成立。

$\widehat{}$证明。首先，我们证明 ψ的随机等度连续性：

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\left[\sup_{|s-t|\le\delta}\left|\widehat{\psi}(s)-\widehat{\psi}(t)\right|\ge\varepsilon\right]=0,$$

对每个 $\varepsilon>0$。固定 $\delta>0$并记

$$w(\widehat{\psi};\delta)=\sup_{|s-t|\le\delta}\left|\widehat{\psi}(s)-\widehat{\psi}(t)\right|.$$

对于 $a\in\{0,1\}$，令 $\Delta_a(s,t)=\widehat{\mu}_a(s,X;\,d)-\widehat{\mu}_a(t,X;\,d)$和 $R(s,t)=\phi(s;\,\mathcal{D}_d)-\phi(t;\,\mathcal{D}_d)-\{\widehat{\mu}_A(s,X;d)-\widehat{\mu}_A(t,X;d)\}$. 那么

$$\widehat{\psi}(s)-\widehat{\psi}(t)=\Delta_1(s,t)-\Delta_0(s,t)+\left\{\tfrac{A}{\widehat{\pi}(X)}-\tfrac{1-A}{1-\widehat{\pi}(X)}\right\}R(s,t).$$

根据假设（A5），$\mathbb{P}\left[\sup_{|s-t|\le\delta}|\Delta_a(s,t)|\right]\le L_{\hat{\mu}}\delta,\ \forall a\in\mathcal{A}$。由于不考虑寿命无限的持久对，我们有 $\|\phi\|_\infty<\infty$。此外，根据引理 2.1，$\phi(t)$关于 t是 1-利普希茨连续的，因此有 $|\phi(s;\mathcal{D}_d)-\phi(t;\mathcal{D}_d)|\le|s-t|$。因此可得$\mathbb{P}\left[\sup_{|s-t|\le\delta}|R(s,t)|\right]\le(L_{\hat{\mu}}+1)\delta$。所以，根据马尔可夫不等式，对于任意固定的 $\varepsilon>0$，有

$$\mathbb{P}\left[w(\widehat{\psi};\delta)\ge\varepsilon\right]\le\frac{\mathbb{P}\left[w(\widehat{\psi};\delta)\right]}{\varepsilon}\lesssim\frac{(3L_{\hat{\mu}}+1)\delta}{\varepsilon},$$

其中，第二个不等式由三角不等式和假设（A1）得出。最后，先取$\delta\to 0$，再取 $\limsup_{n\to\infty}$，得到 $\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}[w(\widehat{\psi};\,\delta)\ge\varepsilon]=0$，这对每个 $\varepsilon>0$，$\widehat{}$均成立，正如所需。既然已经证明了 ψ的随机等度连续性，结论便可直接由引理 C.2 和 Billingsley（1999）的定理 7.5 得出。

□

C.4 假设（A5）下定理 5.2 的证明

证明。 C.3 节中的有限维收敛论证保持不变，因此只需证明随机等度连续性。

固定 $\delta>0$，并记

$$\sup_{|s-t|\le\delta}\left|\widehat{\psi}(s)-\widehat{\psi}(t)\right|\le\mathbb{P}_n\left[\sup_{|s-t|\le\delta}|\widehat{\varphi}_d(s,Z)-\widehat{\varphi}_d(t,Z)|\right].$$

将 $\widehat{\varphi}_d(t,Z)=\varphi_d(t,Z;\,\eta)+\{\widehat{\varphi}_d(t,Z)-\varphi_d(t,Z;\,\eta)\}$ 分解，得到

$$\sup_{|s-t|\le\delta}|\widehat{\varphi}_d(s,Z)-\widehat{\varphi}_d(t,Z)|\le\sup_{|s-t|\le\delta}|\varphi_d(s,Z;\eta)-\varphi_d(t,Z;\eta)|+2\sup_{u\in\mathbb{T}}|\widehat{\varphi}_d(u,Z)-\varphi_d(u,Z;\eta)|.$$

对于第一项，假设（A5）意味着 $t\mapsto\mu_a(t,\ x;d)$ 关于 x 一致满足利普希茨条件，而引理 2.1 给出了 $t\mapsto\phi_{i,\ d}(t)$ 的利普希茨连续性。再结合正值性所蕴含的 $1/\pi(X)$ 和$1/\{1-\pi(X)\}$ 的有界性，可得一个常数 $L_\varphi<\infty$，使得

$$\sup_{|s-t|\le\delta}|\varphi_d(s,Z;\eta)-\varphi_d(t,Z;\eta)|\le L_\varphi\,\delta\quad\text{a.s.}$$

对于第二项，假设（A5）中的一致一致性以及有界分母（假设（A1））意味着

$$\Delta_n:=\sup_{u\in\mathbb{T}}|\widehat{\varphi}_d(u,Z)-\varphi_d(u,Z;\eta)|=o_{\mathbb{P}}(1).$$

因此，

$$\sup_{|s-t|\le\delta}|\widehat{\varphi}_d(s,Z)-\widehat{\varphi}_d(t,Z)|\le L_\varphi\,\delta+2\Delta_n.$$

选择 $\delta = \delta_n \to 0$使得 $L_\varphi \delta_n \le \varepsilon/2$。然后

$$\mathbb{P}\left(\sup_{|s-t|\le\delta_n} \left|\widehat{\psi}(s) - \widehat{\psi}(t)\right| > \varepsilon \right) \le \mathbb{P}(2\Delta_n > \varepsilon/2) \to 0,$$

这便建立了随机等度连续性。其余的紧性条件和结论可参照 C.3 节得出，从而完成假设（A5）下的证明。 □

C.3 节中的有限维收敛论证保持不变，因此只需证明过程 $\widehat{\psi}(t) - \psi_d(t)\}$ 在 $\ell^\infty(\mathbb{T})$ 中的随机等度连续性。

下文中，回顾所提出的估计量为

$$\widehat{\psi}(t) = \mathbb{P}_n\{\widehat{\varphi}_d(t, Z)\}, \qquad \widehat{\varphi}_d(t, Z) \equiv \varphi_d(t, Z; \widehat{\eta}),$$

其中，$\widehat{\eta} = (\widehat{\pi},\ \widehat{\mu}_0,\ \widehat{\mu}_1)$ 是基于来自 $\mathbb{P}$ 的独立样本拟合的，其中根据假设（A2），$\mathbb{P}_n \perp\!\!\!\perp \mathbb{P}$。我们还将 $\varphi_d(t,\ Z) \equiv \varphi_d(t,\ Z;\ \eta)$ 记为 $\eta = (\pi,\ \mu_0,\ \mu_1)$ 对应的真实有效影响函数。

固定 $\delta > 0$。由于 $\psi(t) = \mathbb{P}_n\{\widehat{\varphi}_d(t,\ Z)\}$，我们有

$$\begin{aligned}\sup_{|s-t|\le\delta} \left|\widehat{\psi}(s) - \widehat{\psi}(t)\right| &= \sup_{|s-t|\le\delta} \left|\mathbb{P}_n\left\{\widehat{\varphi}_d(s, Z) - \widehat{\varphi}_d(t, Z)\right\}\right| \\ &\le \sup_{|s-t|\le\delta} \mathbb{P}_n\left|\widehat{\varphi}_d(s, Z) - \widehat{\varphi}_d(t, Z)\right| \le \mathbb{P}_n\Big[\sup_{|s-t|\le\delta} \left|\widehat{\varphi}_d(s, Z) - \widehat{\varphi}_d(t, Z)\right|\Big].\end{aligned} \tag{13}$$

接下来，对于任意 $t \in \mathbb{T}$，

$$\widehat{\varphi}_d(t, Z) = \varphi_d(t, Z) + \left\{\widehat{\varphi}_d(t, Z) - \varphi_d(t, Z)\right\},$$

因此，$\forall s,\ t \in \mathbb{T}$，

$$\begin{aligned}\left|\widehat{\varphi}_d(s, Z) - \widehat{\varphi}_d(t, Z)\right| &\le \left|\varphi_d(s, Z) - \varphi_d(t, Z)\right| + \left|\widehat{\varphi}_d(s, Z) - \varphi_d(s, Z)\right| + \left|\widehat{\varphi}_d(t, Z) - \varphi_d(t, Z)\right| \\ &\le \left|\varphi_d(s, Z) - \varphi_d(t, Z)\right| + 2\sup_{u\in\mathbb{T}}\left|\widehat{\varphi}_d(u, Z) - \varphi_d(u, Z)\right|.\end{aligned} \tag{14}$$

回顾

$$\varphi_d(t, Z) = \mu_1(t, X; d) - \mu_0(t, X; d) + \left\{ \frac{A}{\pi(X)} - \frac{1-A}{1-\pi(X)} \right\} \{\phi_{i,d}(t) - \mu_A(t, X; d)\}.$$

根据假设（A5），对于每个 $a \in \{0, 1\},\ t \mapsto \mu_a(t,\ x;\ d)$在 $\mathbb{T}$ 上关于 x一致地满足利普希茨条件，因此存在 $L_\mu < \infty$使得

$$\sup_{x\in\mathcal{X}} \left|\mu_a(s, x; d) - \mu_a(t, x; d)\right| \le L_\mu |s - t|.$$

根据引理 2.1，$t \mapsto \phi_{i,\ d}(t)$是 1-利普希茨的，因此$|\phi_{i,\ d}(s) - \phi_{i,\ d}(t)| \le |s - t|$。此外，其他位置陈述的正值性条件意味着 $\|1/\pi\|_\infty < \infty$ 和 $\|1/(1-\pi)\|_\infty < \infty$；设

$$C_\pi := \left\| \frac{1}{\pi} \right\|_\infty + \left\| \frac{1}{1-\pi} \right\|_\infty < \infty.$$

那么，对于任意 $s,\ t \in \mathbb{T}$，

$$\begin{aligned}|\varphi_d(s, Z) - \varphi_d(t, Z)| &\le |\mu_1(s, X; d) - \mu_1(t, X; d)| + |\mu_0(s, X; d) - \mu_0(t, X; d)| \\ &\quad + \left| \frac{A}{\pi(X)} - \frac{1-A}{1-\pi(X)} \right| \Big(|\phi_{i,d}(s) - \phi_{i,d}(t)| + |\mu_A(s, X; d) - \mu_A(t, X; d)| \Big) \\ &\le 2L_\mu |s - t| + C_\pi \{\, |s - t| + L_\mu |s - t| \,\} \;=\; L_\varphi\, |s - t|,\end{aligned}$$

其中 $L_\varphi := 2L_\mu + C_\pi(1 + L_\mu) < \infty$。因此，

$$\sup_{|s-t|\le\delta} |\varphi_d(s, Z) - \varphi_d(t, Z)| \le L_\varphi\, \delta \qquad \text{a.s.} \tag{15}$$

由公式 14，定义

$$\Delta_n(Z) \coloneqq \sup_{u\in\mathbb{T}} \big|\widehat{\varphi}_d(u,Z) - \varphi_d(u,Z)\big|.$$

注意

$$\begin{aligned}\widehat{\varphi}_d(t,Z) - \varphi_d(t,Z) = {} & \{\hat{\mu}_1(t,X;d) - \mu_1(t,X;d)\} - \{\hat{\mu}_0(t,X;d) - \mu_0(t,X;d)\} \\ & + \Big(\frac{A}{\hat{\pi}(X)} - \frac{A}{\pi(X)}\Big)\{\phi_{i,d}(t) - \hat{\mu}_1(t,X;d)\} \\ & - \Big(\frac{1-A}{1-\hat{\pi}(X)} - \frac{1-A}{1-\pi(X)}\Big)\{\phi_{i,d}(t) - \hat{\mu}_0(t,X;d)\} \\ & + \left\{\frac{A}{\pi(X)} - \frac{1-A}{1-\pi(X)}\right\}\big\{\mu_A(t,X;d) - \hat{\mu}_A(t,X;d)\big\}.\end{aligned}$$

现在，取 $\sup_{t\in\mathbb{T}} |\cdot|$ 并使用三角不等式。假设（A1）为 $\hat{}\pi$ 给出有界分母，正值性条件为 π 给出有界分母。此外，我们有 $\|\phi_{i,d}\|_\infty < \infty$。因此，存在一个几乎必然有限的随机常数 $C(Z)$（仅依赖于 $\|\phi_{i,d}\|_\infty$、$\|1/\pi\|_\infty$、$\|1/(1-\pi)\|_\infty$、$\|1/\hat{}\pi\|_\infty$ 以及 $\|1/(1-\hat{}\pi)\|_\infty$），使得

$$\Delta_n(Z) = \sup_{t\in\mathbb{T}} |\widehat{\varphi}_d(t,Z) - \varphi_d(t,Z)| \le C(Z)\Big\{ \max_{a\in\{0,1\}} \sup_{t\in\mathbb{T}} |\hat{\mu}_a(t,X;d) - \mu_a(t,X;d)| + |\hat{\pi}(X) - \pi(X)|\Big\}. \tag{16}$$

现在，假设（A5）蕴含关于 (t,x) 的一致一致性 $\sup_{t,x} |\hat{}\mu_a(t,x;d) - \mu_a(t,x;d)| = o_{\mathbb{P}}(1)$ 以及 $\sup_x |\hat{}\pi(x) - \pi(x)| = o_{\mathbb{P}}(1)$，因此方程 16 的右端为 $o_{\mathbb{P}}(1)$。因此，

$$\Delta_n(Z) = o_{\mathbb{P}}(1). \tag{17}$$

结合公式 14、公式 15 和公式 17 可得

$$\sup_{|s-t|\le\delta} \big|\widehat{\varphi}_d(s,Z) - \widehat{\varphi}_d(t,Z)\big| \le L_\varphi\,\delta + 2\Delta_n(Z).$$

代入公式 13 得

$$\sup_{|s-t|\le\delta} \big|\widehat{\psi}(s) - \widehat{\psi}(t)\big| \le L_\varphi\,\delta + 2\,\mathbb{P}_n\{\Delta_n(Z)\}.$$

由于根据公式 17，$\mathbb{P}_n\{\Delta_n(Z)\} \to 0$ 依概率成立，对于任意 $\varepsilon > 0$，取一个确定性序列 $\delta_n \to 0$，使得 $L_\varphi\delta_n \le \varepsilon/2$。于是

$$\mathbb{P}\left(\sup_{|s-t|\le\delta_n} \big|\widehat{\psi}(s) - \widehat{\psi}(t)\big| > \varepsilon\right) \le \mathbb{P}(2\,\mathbb{P}_n\{\Delta_n(Z)\} > \varepsilon/2) \to 0,$$

这验证了随机等度连续性，从而完成了在假设 (A5) 下定理 5.2 的证明。

C.5 定理 5.3 的证明

我们首先引入若干定义，这些定义将在推导稳定性结果时使用 s.

定义 C.1 (匹配). 令 $\triangle$ 表示持久图的对角线。两个持久图 $\mathcal{D}$ 和 $\mathcal{D}'$ 之间的匹配 定义为一个子集 $m \subset \mathcal{D} \times \mathcal{D}'$，使得 $\mathcal{D} \setminus \triangle$ 和 $\mathcal{D}' \setminus \triangle$ 中的每个点在 m 中恰好出现一次。

定义 C.2（Wasserstein 距离）。 两个持久图 q-th Wasserstein 距离 $\mathcal{D}$ 和 $\mathcal{D}'$ 为

$$W_q(\mathcal{D},\mathcal{D}') = \inf_{\text{matching } m} \left(\sum_{(p,p')\in m} \|p - p'\|_\infty^q\right)^{1/q}.$$

我们将满足定义 C.2 的匹配 m 称为最优 W_q 匹配。

设 $\phi = \phi(t;\ \mathcal{D},\ r)$ 和 $\phi' = \phi(t;\ \mathcal{D}',\ r)$ 分别为持久图 $\mathcal{D}$ 和 $\mathcal{D}'$ 对应的幂加权轮廓。给定 $p = (a_p,\ b_p) \in \mathcal{D}$ 和 $p' = (a_{p'},\ b_{p'}) \in \mathcal{D}'$，我们将分别用 $w_p = |b_p - a_p|^r$ 和 $w_{p'} = |b_{p'} - a_{p'}|^r$ 表示 p 和 p' 对应的幂权重，其中 $r > 0$。注意，在假设（A6）的图有界性条件下，幂权重中的绝对值可以省略。因此，在本节中，我们始终假设 $w_p = (b_p - a_p)^r$ 和 $w_{p'} = (b_{p'} - a_{p'})^r$。

现在我们建立以下引理，它是后续主要定理推导所必需的预备结果。

引理 C.3. 给定一个匹配 $m \subset \mathcal{D} \times \mathcal{D}'$，对于任意 $(p, p') \in m$，

$$|w_p - w_{p'}| \ \le\ 2rc_{pp'}^{r-1} \cdot \|p - p'\|_\infty,$$

其中 $c_{pp'} \in (\min\{\, b_p - a_p,\ b_{p'} - a_{p'} \}, \max\{\, b_p - a_p,\ b_{p'} - a_{p'} \})$ 是一个正数常量，它满足函数 $g(x) = x^r, x \in [0, \infty)$ $r > 0$在给定点 p和p'处的中值定理。因此，

$$|w_p - w_{p'}| \ \le\ 2rc^{r-1} \cdot \|p - p'\|_\infty,$$

for

$$c \ =\ \max_{(p,p') \in m} \ \max\{\, b_p - a_p,\ b_{p'} - a_{p'} \} \ =\ \max\{\, b_x - a_x : x \in \mathcal{D} \cup \mathcal{D}' \},$$

所以 c 仅取决于加权指数 r （通过该界）以及$\mathcal{D}$ 和 $\mathcal{D}'$中最大的寿命。

证明。

$$\begin{aligned}
|w_p - w_{p'}| &= |(b_p - a_p)^r - (b_{p'} - a_{p'})^r| \\
&= |r\, c_{pp'}^{r-1}\, \{(b_p - a_p) - (b_{p'} - a_{p'})\}| \\
&= r\, c_{pp'}^{r-1}\, |(b_p - a_p) - (b_{p'} - a_{p'})| \\
&\le r\, c_{pp'}^{r-1}\, \{|a_p - a_{p'}| + |b_p - b_{p'}|\} \\
&\le 2r\, c_{pp'}^{r-1}\, \max\{|a_p - a_{p'}|, |b_p - b_{p'}|\} \\
&= 2r\, c_{pp'}^{r-1}\, \|p - p'\|_\infty,
\end{aligned}$$

其中第二个等式用到了中值定理。 □

注意，当 $r = 1$ 时，该界不涉及 c（参见定理中的特殊情形）。

现在我们给出定理 5.3 的证明。

Proof. Let $w_p = (b_p - a_p)^r$, $w_{p'} = (b_{p'} - a_{p'})^r$, and define

$$S = \sum_{p \in \mathcal{D}} w_p, \qquad S' = \sum_{p' \in \mathcal{D}'} w_{p'}, \qquad \phi = \frac{1}{S} \sum_{p \in \mathcal{D}} w_p \Lambda_p, \qquad \phi' = \frac{1}{S'} \sum_{p' \in \mathcal{D}'} w_{p'} \Lambda_{p'},$$

其中， Λ_p 表示以 p 为中心的帐篷函数，使得 $\|\Lambda_p\|_\infty = (b_p - a_p)/2$。如有需要，在图中补上对角线后，最优 W_1 匹配 m^* 是 $\mathcal{D}$ 与 $\mathcal{D}'$之间的双射，因此

$$\|\phi - \phi'\|_\infty \ =\ \left\| \sum_{(p,p') \in m^*} \Big(\frac{w_p \Lambda_p}{S} - \frac{w_{p'} \Lambda_{p'}}{S'} \Big) \right\|_\infty \ \le\ \sum_{(p,p') \in m^*} \left\| \frac{w_p \Lambda_p}{S} - \frac{w_{p'} \Lambda_{p'}}{S'} \right\|_\infty .$$

将每个被加项拆分为

$$\left\| \frac{w_p \Lambda_p}{S} - \frac{w_{p'} \Lambda_{p'}}{S'} \right\|_\infty \ \le\ \left\| \frac{w_p}{S} (\Lambda_p - \Lambda_{p'}) \right\|_\infty + \|\Lambda_{p'}\|_\infty \left| \frac{w_p}{S} - \frac{w_{p'}}{S'} \right| .$$

对 $(p,\ p')$ 求和后， $\in m^*$ 得到

$$\|\phi - \phi'\|_\infty \ \le\ T_1 + T_2, \qquad T_1 := \frac{1}{S} \sum_{(p,p') \in m^*} w_p \|\Lambda_p - \Lambda_{p'}\|_\infty, \quad T_2 := \sum_{(p,p') \in m^*} \|\Lambda_{p'}\|_\infty \left| \frac{w_p}{S} - \frac{w_{p'}}{S'} \right| .$$

对 T_1 的界。 根据帐篷函数的1-Lipschitz性质，我们有 $\|\Lambda_p - \Lambda_{p'}\|_\infty \le \|p - p'\|_\infty$。因此，

$$T_1 \;\le\; \frac{1}{S}\sum_{(p,p')\in m^*} w_p \|p-p'\|_\infty \;\le\; \frac{1}{S}\Big(\sum_{p\in\mathcal{D}} w_p\Big)\Big(\sum_{(p,p')\in m^*}\|p-p'\|_\infty\Big) \;=\; W_1(\mathcal{D},\mathcal{D}'),$$

其中第二个不等式由霍尔德不等式的一个简单推论得出：

$$\sum_i \beta_i x_i \;\le\; \|\beta\|_1 \, \|x\|_\infty \;\le\; \|\beta\|_1 \, \|x\|_1 \;=\; \Big(\sum_i \beta_i\Big)\Big(\sum_i x_i\Big),$$

对于 $\beta_i, x_i \ge 0$，最后一个等式是 m^* 的 W_1 代价的定义。

关于 T_2的界。 观察代数恒等式

$$\Big|\frac{w_p}{S} - \frac{w_{p'}}{S'}\Big| = \frac{|(S'-S)w_p + S(w_p - w_{p'})|}{SS'} \;\le\; \frac{|S'-S|\, w_p}{SS'} + \frac{|w_p - w_{p'}|}{S'}.$$

因此 $T_2 \le T_{2a} + T_{2b}$，其中

$$T_{2a} := \sum_{(p,p')\in m^*} \|\Lambda_{p'}\|_\infty \frac{|S'-S|\, w_p}{SS'}, \qquad T_{2b} := \sum_{(p,p')\in m^*} \|\Lambda_{p'}\|_\infty \frac{|w_p - w_{p'}|}{S'}.$$

我们首先处理项 T_{2a}。利用霍尔德不等式分离求和并消去 S，我们得到

$$T_{2a} \;\le\; \Big(\sum_{p'\in\mathcal{D}'} \|\Lambda_{p'}\|_\infty\Big)\frac{|S'-S|}{S'}.$$

由于$S' - S = \sum_{(p,p')\in m^*}(w_{p'} - w_p)$，由三角不等式和引理 C.3 可得

$$|S'-S| \;\le \sum_{(p,p')\in m^*} |w_p - w_{p'}| \;\le \sum_{(p,p')\in m^*} 2r\, c_{pp'}^{r-1}\, \|p-p'\|_\infty \;\le\; 2r\, c^{r-1} \sum_{(p,p')\in m^*} \|p-p'\|_\infty.$$

此外，$\sum_{p'\in\mathcal{D}'} \|\Lambda_{p'}\|_\infty = \frac{1}{2}\sum_{p'\in\mathcal{D}'}(b_{p'} - a_{p'})$ 与 $S' = \sum_{p'\in\mathcal{D}'}(b_{p'} - a_{p'})^r$，故由假设（A6）中的有界性条件，

$$\frac{\sum_{p'\in\mathcal{D}'} \|\Lambda_{p'}\|_\infty}{S'} = \frac{1}{2}\cdot\frac{\sum_{p'}(b_{p'} - a_{p'})}{\sum_{p'}(b_{p'} - a_{p'})^r} \;\le\; \frac{L}{2}.$$

因此

$$T_{2a} \;\le\; \frac{L}{2}\, 2r\, c^{r-1} \sum_{(p,p')\in m^*} \|p-p'\|_\infty \;=\; L\, r\, c^{r-1}\, W_1(\mathcal{D},\mathcal{D}').$$

对 T_{2b} 的分析遵循与上文相同的步骤。再次由引理 C.3 以及假设（A6）中的比率有界性条件，

$$T_{2b} \;\le\; \frac{\sum_{p'\in\mathcal{D}'} \|\Lambda_{p'}\|_\infty}{S'} \sum_{(p,p')\in m^*} |w_p - w'_{p'}| \;\le\; \frac{L}{2} \sum_{(p,p')\in m^*} 2r\, c_{pp'}^{r-1}\, \|p-p'\|_\infty \;\le\; L\, r\, c^{r-1}\, W_1(\mathcal{D},\mathcal{D}').$$

合并 T_{2a} 与 T_{2b} 的界可得

$$T_2 \;\le\; 2L\, r\, c^{r-1}\, W_1(\mathcal{D},\mathcal{D}').$$

综合 T_1 与 T_2的界，我们得到

$$\|\phi - \phi'\|_\infty \;\le\; \big(1 + 2L\, r\, c^{r-1}\big)\, W_1(\mathcal{D},\mathcal{D}').$$

特殊情形 $r = 1$**.** 当 $r = 1$ 时，引理 C.3 给出 $|w_p - w_{p'}| \le 2\|p - p'\|_\infty$ and

$$\frac{\sum_{p'\in\mathcal{D}'} \|\Lambda_{p'}\|_\infty}{S'} = \frac{\sum_{p'}(b_{p'} - a_{p'})/2}{\sum_{p'}(b_{p'} - q_{p'})} = \frac{1}{2},$$

因此，T_{2a} 和 T_{2b} 各自均由 $W_1(\mathcal{D},\mathcal{D}')$界定，而 $T_1 \le W_1(\mathcal{D},\mathcal{D}')$，由此得到

$$\|\phi - \phi'\|_\infty \le 3\, W_1(\mathcal{D},\mathcal{D}').$$

□

表1： L_1 平均估计值与真实效应之间的距离，以及每个估计量在20次重复中的标准差。 $\mathcal{H}_d$ 表示同调维数 d，SARS-CoV-2和ORBIT的尺度为1e-05，GEOM-Drugs的尺度为1e-02。

	SARS-CoV-2		**GEOM-Drugs**				**ORBIT**			
	$\mathcal{H}_0$		$\mathcal{H}_0$		$\mathcal{H}_1$		$\mathcal{H}_0$		$\mathcal{H}_1$	
	L_1 分布	Std.	L_1 分布	Std.	L_1 分布	Std.	L_1 分布	Std.	L_1 分布	Std.
PI	4.544	2.712	3.215	0.740	7.041	1.283	0.571	0.530	36.063	37.704
IPW	2.936	9.035	**1.670**	5.756	18.886	8.337	**0.073**	3.722	18.210	93.978
AIPW	**1.246**	3.099	3.296	1.470	**2.136**	1.553	0.121	0.802	**6.920**	47.127

C.6 推论 5.4 的证明

我们仅给出简要概述，因为该论证可由标准经验过程方法$\widehat{}$（例如，Van Der Vaart & Wellner, 1996年）推出。根据定理 5.2， $\sqrt{n}(\psi_{\mathrm{AIPW},d} - \psi_d) \rightsquigarrow \mathbb{G}_d$ 属于 $\ell^\infty(\mathbb{T})$，而将连续映射定理应用于 $f \mapsto \|f\|_\infty$ 可推出，在 H_0下， $T_n \rightsquigarrow \|\mathbb{G}_d\|_\infty$ 。如果乘子自助法在给定数据的条件下能一致地逼近 $\|\mathbb{G}_d\|_\infty$的分布，那么其 $(1-\alpha)$ 分位数 $c_{1-\alpha}$ 将给出渐近有效的水平为-α 的检验；在任何满足 $\|\psi_d\|_\infty > 0$的固定备择下，我们有 $T_n \geq \sqrt{n}\|\psi_d\|_\infty + o_{\mathbb{P}}(\sqrt{n}) \to \infty$，因此检验功效趋于 1。

D 实验详情

我们在三个基准数据集上进行实验：SARS-CoV-2 CT扫描图像数据集、GEOM-Drugs分子图数据集和ORBIT点云数据集。在所有实验中，我们构建一个合成反事实数据集 $\{(X_i, A_i, Y_i^0, Y_i^1)\}_{i=1}^n$，以便根据已知的真实效应评估估计量。首先，我们随机选择两个数据样本并将其配对，形成每个潜在结果对；随后以一种能够产生明确拓扑对比的方式，将其中一个分配给 Y^0，另一个分配给 Y^1。接下来，我们根据随机数据生成过程和处理分配机制生成协变量 X 和处理 A。所有实验均采用完全相同的流程，具体细节如下。

数据生成过程。我们假设协变量 $X \in \mathbb{R}^5$具有亚组结构。协变量由两个子组生成，每个子组均服从具有不同均值向量 μ_1， μ_2和共同协方差矩阵 Σ的多元高斯分布。协方差矩阵 Σ 设为对角线元素为 0.5 的对角矩阵，每个子组的均值向量分别设为 $\mu_1 = [1, 0.6, -0.7, 2.2, -1]^T$和 $\mu_2 = [0.4, -0.4, -0.6, 3.3, 3]^T$。一半样本的协变量从 $N(\mu_1, \Sigma)$ 生成，其余样本则从 $N(\mu_2, \Sigma)$ 生成。

处理机制。给定协变量 X，处理 $A \in \{0, 1\}$ 以概率$\pi(X) = expit(-0.5X_1 - 0.1X_2 + 0.6X_3 + 0.1X_4 + 0.1X_5 + 0.5X_2X_3 - 0.7X_1X_3)$分配。该处理机制经过设计，使一个子组接受处理的概率高于另一个子组（见图 6a）。分配处理后，观测结果为 $Y = AY^1 + (1-A)Y^0$。

上述过程重复20次，以生成 X 与 A具有不同实现的多个数据集；对每个数据集，分别计算插件估计、逆概率加权和增广逆概率加权估计量。这些估计量的构造方式为：采用傅里叶基展开的函数对标量回归对轮廓回归函数 μ_a 进行建模，并使用随机森林分类器估计倾向评分 π。为评估各估计量的表现，我们考察在20次运行中计算得到的逐点均值以及逐点1个标准差误差带。此外，我们在表1中报告平均估计函数与真实拓扑效应函数之间的 L_1 距离，作为估计量准确性的单一数值摘要。为完整起见，我们还在表1中给出各估计量的标准差：先对经验协方差矩阵取平均以得到变异性的标量摘要，再取其平方根。

表2：计算每个数据集加权轮廓所需的运行时间（秒）。

数据（样本量）	**SARS-CoV-2（2481）**	**GEOM-Drugs（30433）**	**ORBIT（2000）**
运行时间	244 秒	21 秒	25 秒

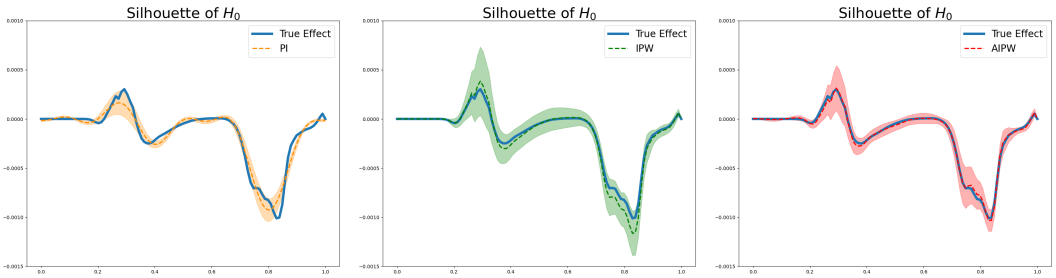


图5：SARS-CoV-2 数据集上 PI、逆概率加权和增广逆概率加权估计量的逐点均值（虚线）及逐点 1 个标准差误差带（阴影区域）的可视化。真实拓扑因果效应以蓝线表示。

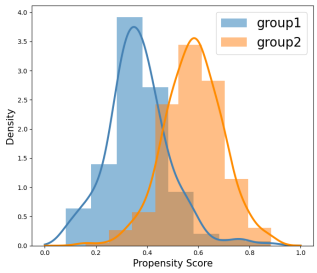
D.1 实现注意事项

在讨论各实验的细节之前，我们首先概述与实施本方法相关的若干实际考量。

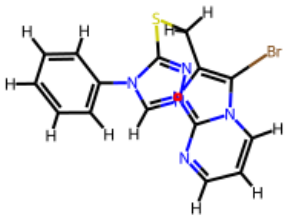
样本分割。 如第4节所述，样本分割支持使用任意复杂的干扰参数估计量。因此，采用样本分割对于正确实施至关重要，下面说明其在实践中的具体操作方法。令 $D = \{Z_i = (X_i, A_i, Y_i)\}_{i=1}^n$ 表示一个大小为 n 的独立同分布观测样本。我们将 D 随机分割为两个独立样本 D_1 和 D_2，每个样本的大小均为 $n/2$。首先，我们使用 D_1 拟合干扰参数估计量 $\hat{}\eta$，并使用 D_2 计算 ψ，以估计 ψ。然后，我们交换两个样本，使用 D_2 进行训练，并使用 D_1 进行估计。如果我们对所得的两个估计量 ψ取平均，就可以恢复全样本效率。

计算成本。 我们的框架依赖于持久同调，这会产生较高的计算成本；持久同调的标准算法具有 $O(N^3)$ 的复杂度（Edelsbrunner 与 Morozov），其中 N 表示单纯形的数量。为了部分降低这种计算开销并提高可扩展性，人们已经开发出多种算法技术。例如，Chen 与 Kerber（2011）提出了一种用于持久同调的输出敏感算法，其复杂度为 $O(N^2 \log^3 N \log\log N)$。对于许多只需 0维持久同调的应用（例如我们的 SARS-CoV-2 实验），还可以通过采用高效算法进一步提升可扩展性；这些算法的运行时间为 $O(N \log N)$（Edelsbrunner 与 Morozov）。然而，即使采用改进后的算法，我们利用持久同调的方法仍可能具有较高的计算负担，并且在大规模数据集上的可扩展性受到限制，因为这一局限性是持久同调本身所固有的。为了实现更高的可扩展性，可以考虑将我们的框架扩展到基于高效欧拉示性曲线的拓扑描述符，尽管基于欧拉示性曲线的描述符通常不具备基于持久同调的方法所具有的表达能力。

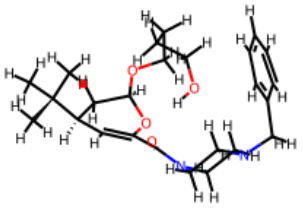
为了说明我们的方法对于规模并不过大的数据集完全可行，我们报告了在本研究使用的所有数据集上计算加权轮廓的实测运行时间。对于 SARS-COV-2 数据集，我们在 1252 名感染患者和 1229 名未感染患者的 CT 图像上计算加权轮廓。GEOM-Drugs 使用完整数据集中的 30422 个图进行计算。对于 ORBIT 数据集，我们对 1000 对 (Y^1, Y^0) 计算加权轮廓，总计得到 2,000 个轮廓函数。轮廓函数的计算使用 GUDHI（GUDHI 项目，2021）库，并在 Apple M2 处理器上运行；运行时间结果见表2。



(a) 倾向评分密度，按子组划分（第2组更可能接受处理）。



(b) 样本分子图 A（Axelrod & Gomez-Bombarelli, 2022）。



（c）示例分子图 B（Axelrod & Gomez-Bombarelli，2022）。

图6：（a）按子组划分的倾向评分分布；（b–c）代表性分子图。

D.2 SARS-CoV-2

SARS-CoV-2 数据集包含从巴西真实患者收集的 CT扫描，这些患者感染或未感染 COVID-19。在我们的实验中，按照以下步骤构造 500 对潜在结果对 (Y^0, Y^1)：(i) 抽取 500 张感染图像并分配给 Y^0；(ii) 抽取 375 张未感染图像和 125 张感染图像并分配给 Y^1；(iii) 将 Y^0 andY^1 随机配对。根据实验设计，拓扑平均处理效应表现出处理效应，因为 Y^1 中有 75% 为未感染者，而 Y^0 中的每个人都已感染。轮廓函数采用子水平集滤过，在带有 $r = 0.1$ 的滤过立方复形上计算（见附录A）。对于干扰参数估计量，函数对标量回归采用 10 个基函数实现，而随机森林则由 100 棵树构成。各估计量逐点的 1 个标准差误差带见图5，估计量性能的数值汇总见表1。

D.3 GEOM-DRUGS

GEOM-Drugs数据集包含分子化合物的图结构表示，如图6b和6c所示。所有图节点都有六个特征，我们将每个节点的标量权重设为这些特征值的随机凸组合。每条边的权重设为其相邻节点权重中的最大值。由于一维同调特征的死亡时间为无穷大，我们将其替换为从 $[6.5, , 8.5]$ 中均匀采样的值。在我们的实验中，按照以下步骤构造了 1000 对潜在结果（Y^0, Y^1）：(i) 随机抽取 1000 个含有一个环的图，并将其分配给 Y^0；(ii) 随机抽取 750 个含有两个环的图和 250 个含有一个环的图，并将其分配给 Y^1。这种分配方案使反事实组之间产生了显著的拓扑差异。我们使用 $r = 1$ 计算轮廓函数，而 μ_a 和 π 的干扰参数估计量分别采用 5 个基函数和 100 棵树。图7展示了各同调维数下估计量的逐点均值和逐点 1 个标准差误差带；估计量性能的数值汇总见表1。

D.4 ORBIT

ORBIT 数据集（Adams等人，2017）是一个合成点云数据集，通过模拟由参数 s 表征的不同动力系统生成。给定一个随机初始点（x_0， y_0） $\in [0, 1]^2$ 和 $s > 0$，我们按如下方式生成包含 1000 个点的点云：

$$x_{n+1} = x_n + sy_n(1 - y_n) \mod 1$$
$$y_{n+1} = y_n + sx_n(1 - x_n) \mod 1,$$

在本实验中，我们使用 $s = 3.5$，4，4。1 为 s 的每个取值生成 1000 个样本，所得点云如图2a所示：$s = 3.5$ （右上）、 $s = 4$ （下方）、 $s = 4.1$（左上）。对于由 $s = 3.5$，4，4。1 生成的每个点云三元组，我们随机选取两个点云，并以 0。7 的概率将对应较高 s 值的点云指定为 Y^1。这一过程

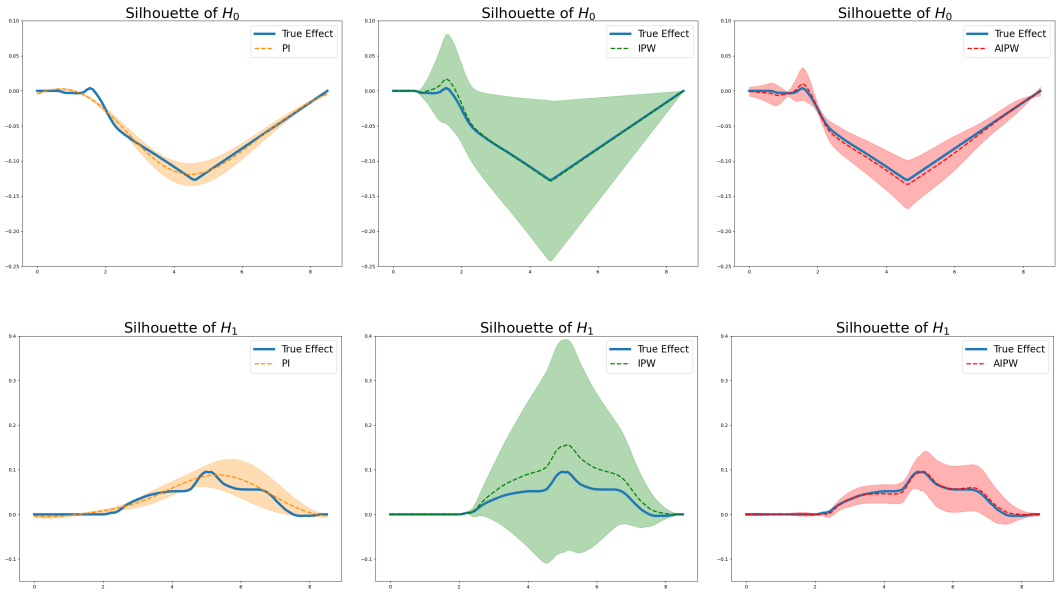


图 7：在 GEOM-Drugs 数据集上，0 维（上）和 1 维（下）的 PI、IPW 和 AIPW 估计量的逐点均值（虚线）及逐点 1 个标准差误差带（阴影区域）的可视化。真实拓扑因果效应以蓝线表示。

为全部 1000 个样本生成了匹配的潜在结果对（Y^0，Y^1），其中处理组的潜在结果经过有意设计，具有更为显著的拓扑特征。这里，我们使用 3 个基函数对轮廓回归函数进行建模，并使用 100 棵树估计倾向评分；轮廓函数则使用 Alpha过滤计算（见附录A），幂权重为 $r = 3$。

结果。 图 8 展示了同调维数 0 和 1 下真实的目标轮廓函数，清楚地表明处理对点云的一阶同调特征存在因果处理效应。具体而言，1 维轮廓函数的正值表明处理组点云中出现了孔洞。我们的目标是恢复该剪影函数的幅度，因为数值越大，表示结构变化越显著。与先前结果一致，逆概率加权估计量往往会高估处理效应，而代入估计量则会低估处理效应。相比之下，增广逆概率加权估计量对真实轮廓函数的估计准确得多。此外，0 维同调特征的轮廓函数接近于零，这表明数据的 0 维同调（包括连通分支）在处理后基本保持不变。图 9 展示了 ORBIT 数据集上 PI、IPW 和 AIPW 各估计量的逐点均值及逐点 1 个标准差误差带；估计量性能的数值汇总见表 1。

假设检验。 ORBIT 数据集为对推论 5.4 中提出的假设检验进行实证评估提供了合适的场景，因为 0 维轮廓函数未显示出明显的拓扑效应，而 1 维轮廓函数则显示出显著的拓扑效应（图 8）。因此，我们进行了原假设为 H_0： $W_1(\mathcal{D}^1,\ \mathcal{D}^0) = 0$的假设检验，以考察检验结果是否与我们对图 8 的视觉解读一致。我们ˆ生成了 1000 个自助法样本 $\mathbb{G}_n$ 并使用 $\alpha = 0.05$。对于 0 维同调，检验统计量为ˆ$T_n = 0.0037$，而 $\|\mathbb{G}_n\|_\infty$ 的 $1-\alpha$ 分位数为 0.0277。因此，与我们的预期一致，我们无法拒绝原假设。对于 1 维同调，检验统计量为 $T_n = 0.4242$，ˆ而 $\|\mathbb{G}_n\|_\infty$ 的 $1-\alpha$ 分位数为 0.1107。正如预期，检验拒绝了原假设，这表明数据的 1 维同调中存在显著的处理诱导的拓扑变化。

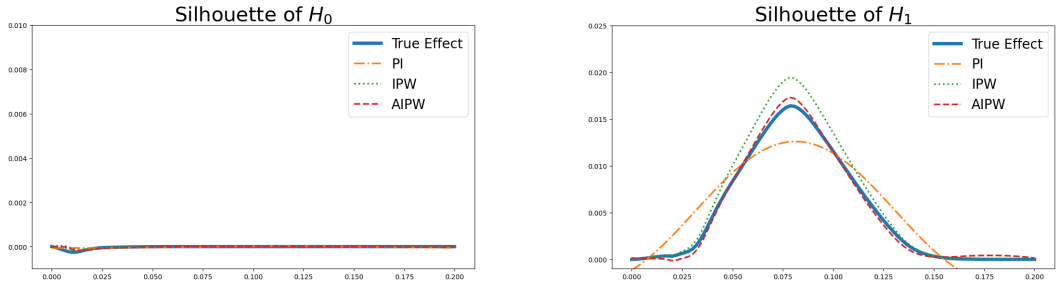


图 8：ORBIT 数据集的 0 维（左）和 1 维（右）真实轮廓函数及其 PI、IPW、AIPW 估计值。

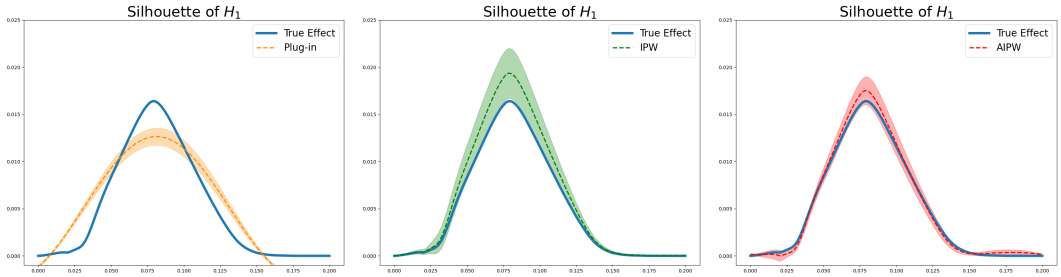
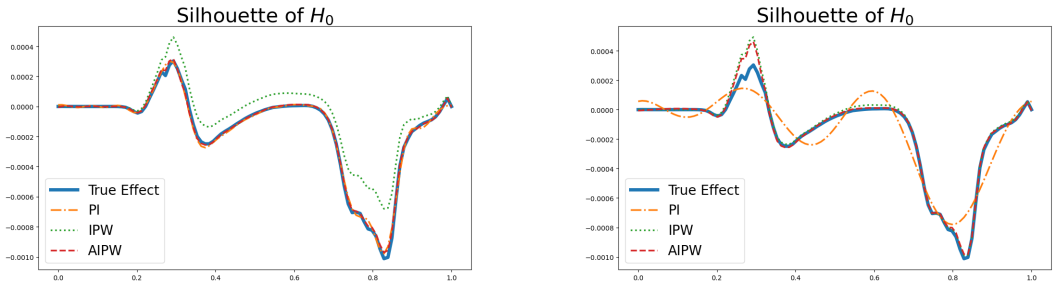


图 9：ORBIT 数据集上 1 维 PI、IPW 和 AIPW 估计量的逐点均值（点线）及逐点 1 个标准差误差带（阴影区域）的可视化结果。真实拓扑因果效应以蓝线表示。

D.5 模型设定错误下的估计

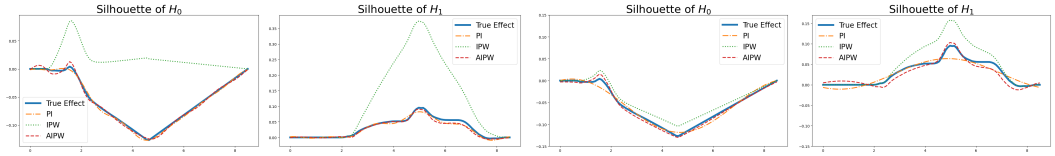
在之前的实验中，我们使用复杂的非参数模型灵活地估计干扰函数。本节中，我们考察估计量在干扰函数模型严重设定错误时的表现。具体而言，我们希望观察估计量在以下情形中的表现差异：(i) π 严重设定错误，而 μ 是正确模型；(ii) μ严重设定错误，而 π 是正确模型。

由于我们了解真实的处理过程，因此将正确的 π 建模为 $\hat{\pi}(X) = expit(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_2 X_3 - \beta_7 X_1 X_3)$，而对于设定错误的模型，则使用 $\hat{\pi}(X) = expit(\beta_1 X_1 + \beta_2 X_3)$。对于结果回归 μ，由于其真实形式未知，我们仍采用基于傅里叶基展开的函数对标量回归。我们选择足够多的基函数来表示设定正确的模型，并刻意减少基函数数量以构建设定错误的模型。具体而言，对于 SARS-CoV-2、GEOM-Drugs 和 ORBIT 数据集，正确的回归模型分别使用 30、10 和 5 个基函数，而设定错误的回归模型分别使用 7、2 和 2 个基函数。图 10、11 和 12 展示了可视化结果，表 3 则提供了估计量性能的定量汇总。我们可以观察到，在所有数据集和模型设定错误情形下，增广逆概率加权估计量通常表现最佳。



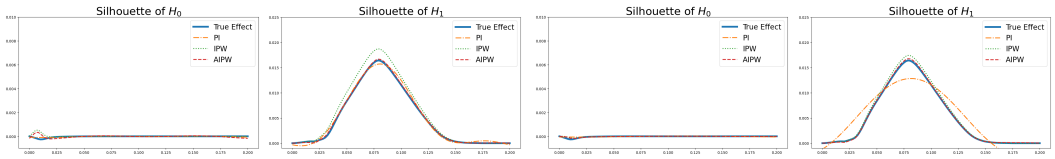
(a) π 设定错误， μ 设定正确　　(b) μ设定错误， π 设定正确

图10：当其中一个干扰函数设定错误时，SARS-CoV-2 数据集上0维同调特征的真实轮廓及其插件估计、逆概率加权和增广逆概率加权估计。



(a) π 设定错误， μ 设定正确　　(b) μ设定错误， π 设定正确

Fig图11：GEOM-Drugs数据集上 0 维和 1 维同调 1
feat特征的真实轮廓函数及其插件估计、逆概率加权和增广逆概率加权估计，其中一个干扰函数设定错误。



(a) π 设定错误， μ 设定正确　　(b) μ设定错误， π 设定正确

图12：ORBIT 数据集上 0 维和 1 维同调特征的真实轮廓函数及其插件估计、逆概率加权和增广逆概率加权估计，其中一个干扰函数设定错误。

表 3： L_1 当任一干扰函数设定错误时，平均估计值与真实效应之间的距离，以及各估计量在 20 次重复实验中的标准差。$\mathcal{H}_d$ 表示同调维数 d，SARS-CoV-2 和 ORBIT 的尺度为 1e-05，GEOM-Drugs 的尺度为 1e-02。

SARS-CoV-2 ($\mathcal{H}_0$)	π **设定错误，** μ **设定正确**		μ **设定错误，** π **是正确的**	
	L_1 分布	Std.	L_1 分布	Std.
PI	1.390	3.132	7.985	3.132
IPW	9.674	4.421	2.243	4.328
AIPW	**0.786**	2.761	**1.529**	3.395

GEOM-Drugs	π **设定有误，** μ **是正确的**				μ **设定有误，** π **设定正确**			
	$\mathcal{H}_0$		$\mathcal{H}_1$		$\mathcal{H}_0$		$\mathcal{H}_1$	
	L_1 分布	Std.	L_1 分布	Std.	L_1 分布	Std.	L_1 分布	Std.
PI	**2.006**	1.133	4.022	0.865	4.376	1.133	6.879	0.865
IPW	61.796	3.602	77.963	4.426	10.299	2.831	16.245	3.564
AIPW	2.343	1.168	**3.432**	0.897	**3.380**	1.390	**6.842**	1.112

ORBIT	π **设定有误，** μ **正确**				μ **设定有误，** π **正确**			
	$\mathcal{H}_0$		$\mathcal{H}_1$		$\mathcal{H}_0$		$\mathcal{H}_1$	
	L_1 分布	Std.	L_1 分布	Std.	L_1 分布	Std.	L_1 分布	Std.
PI	**0.514**	0.688	7.298	37.179	0.600	0.688	36.323	37.179
IPW	0.766	2.300	17.576	45.287	**0.156**	1.482	7.092	55.763
AIPW	1.347	0.681	**2.605**	39.582	0.284	0.748	**2.354**	40.813