基于偏好的条件处理效应与策略学习

Dovid Parnas*以色列理工学院 Mathieu Even*Inria、Inserm、蒙彼利埃大学，法国 Julie Josse Inria、Inserm、蒙彼利埃大学，法国 Uri Shalit 统计与运筹学系，特拉维夫大学

arXiv:2602.03823v1 [stat.ML] 3 Feb 2026

2026年2月4日

摘要

我们提出了一种新的基于偏好的条件处理效应估计与策略学习框架，该框架建立在条件偏好处理效应（CPTE）之上。CPTE 仅要求根据偏好规则对结果进行排序，从而能够灵活地对多变量结果、有序结果或偏好驱动的结果的异质性效应进行建模。该框架统一了条件必要性与充分性概率、条件胜率比以及广义两两比较等应用。尽管基于比较的估计目标具有内在的不可识别性，CPTE 仍提供了可解释的目标，并为以往不可识别的估计目标给出了新的可识别性条件。我们通过匹配、分位数回归和分布回归提出了多种估计策略，并进一步设计了高效的影响函数估计量，以纠正插件偏差并最大化策略价值。合成实验和半合成实验表明，该方法在性能和实际影响方面均有明显提升。

1 引言

在医疗保健和公共政策等高风险领域，临床医生和政策制定者必须决定采用哪些治疗，因为治疗效果在不同个体之间可能存在很大差异。群体层面平均值往往远远不够：对每个人而言，正确的选择取决于其具体特征、风险因素以及可能的反应。要真正实现数据驱动决策的个性化，我们必须估计个体效应，并学习依赖特征的策略。

标准方法会估计条件处理效应，比较各治疗组的条件结果，并据此制定策略。然而，尽管条件平均处理效应（CATE）已成为异质性效应估计和策略学习的默认工具，但它可能产生误导。基于 CATE 的策略依赖平均值，忽略了复杂的治疗反应。例如，重尾分布可能导致基于 CATE 的策略偏向某些治疗，而实际上大多数人会因此受损。以薪资为结果，考虑这样一种干预措施：它让许多低薪人群每月承受小额损失，却为少数高薪人群带来巨额收益。即使 CATE 估计得完全准确，也会建议采取该干预措施，尽管这会降低大多数人的福利。在关键设定中，这种偏差不仅仅是统计假象，还会威胁公平、有效的决策。

此外，对于多个结果而言，定义 CATE 策略并非易事。以心力衰竭研究为例， [44] 死亡、中风和住院均是主要关注的结果。这些结果有着自然的层级：死亡最为严重，其次是中风，再次是住院。难点在于如何将这一层级纳入一个严谨、由数据驱动的决策框架。

*共同贡献。

过去十年来，一种主要方法是胜率比[44], ，即将接受处理的患者与对照组患者配对并依次进行比较。首先评估死亡情况；如果双方均存活，则比较中风情况；如果双方均未发生中风，则评估住院情况。这种设计明确遵循了临床优先级。然而，当结果只能通过比较来评判时，量化处理效应、条件效应以及学习策略仍然极具挑战性 [37, 13, 53]。这些效应无法用条件平均处理效应来刻画。核心难题在于底层估计目标的不可识别性；由于因果推断的根本问题 [19]，无法根据可观测数据对其进行估计。事实上，要比较反事实，必须能够获取联合潜在结果分布。然而，即使在随机研究中，也不可能观测到任何个体同时处于处理组和对照组。

本文提出一个通用框架，用于从随机化数据与观察性数据中学习基于偏好的条件治疗效应和策略。基于偏好的学习直接回答“应优先选择哪种处理？”这一问题。它还将策略治疗效应估计与策略学习扩展到可比较但不可量化的结果。第一个例子是必要性与充分性概率（PNS），这是一个经典因果量[37, 53]。PNS可解释为：对于给定个体，处理后的结果严格优于对照结果的概率，并在总体中取平均。尽管相关方向已被探讨 [57, 58, 48], ，但它缺乏一种可识别的、面向个性化决策的条件构造。

更广泛地说，只要有规则规定在两两比较中哪个结果更受偏好，基于偏好的效应便可自然扩展到复杂的多变量结果。该框架涵盖层级结果情境，例如胜率比和广义两两比较[44, 6]。我们的方法超越这些情形，在此类偏好结构下学习条件效应与最优策略，同时还能容纳更一般的、可能非严格的偏好规则。

相关研究。 关于条件因果推断中的处理效应和策略学习，大多数研究都聚焦于单一的实值或二元结局 [3, 56, 30, 4]。本文提出了一个更广泛的框架，将三条相关研究脉络联系起来：*(i)* 不可识别的因果量及其界限，例如 PNS；*(ii)* 胜率比和净获益等两两比较方法；以及*(iii)* 分布回归和分位数处理效应。

必要性与充分性概率（PNS）是指处理改善结果且改善结果离不开该处理的概率 [37, 53]。它可视为个体层面效应的总体平均效应；Kawakami 等人 [25], 、Wang 和 Li [57]研究了其在子总体中的扩展。Lei [31] 给出了 PNS 的总体层面点估计。据我们所知，目前尚无对可识别的条件 PNS 的明确界定。

由于 PNS 及其条件变体在根本上不可识别，以往大多数研究都侧重于在适当假设下从数据中推导可识别界限 [48, 45, 53]。针对二项分布 [37], 、有序 [21, 33, 32, 11], 、连续 [34],分位数处理效应（QTE） [28, 15]以及条件 QTE [1, 15, 10]，研究者已提出多种界定策略。我们不依赖可识别界限（对于条件量或非二元结果，这类界限可能较宽松 [32, 34],），而是提出并估计一种可解释、可识别的代理指标。

Pocock 等人 [44] 与 Buyse [6] 提出了胜率比 和净获益 ，用于在多个分层 结果上比较两种治疗。这些指标使多个结果能够纳入处理效应的评估与决策。Petit 等人 [43]考虑了针对净获益的最优处理规则。然而，我们更广泛的框架更进一步，通过定义具有分层结果的条件处理效应，实现策略价值优化并发展半参数有效估计量。我们的工作借鉴了 [13] 的形式体系以及非因果概率指数模型（PIMs, [52]）。

我们的估计方法基于最近邻方法 [5]和现成的分位数回归方法 [27, 17, 36]来估计干扰参数。我们的方法不同于分布处理效应方法 [38, 60, 41, 24]：我们使用分布方法作为一种手段来估计

基于偏好的效应。

论文提纲与贡献。我们首先在第2节介绍因果与偏好设定，并在第3节定义我们的目标估计目标。我们区分不可识别的条件个体处理效应（ITE）与新的可识别估计目标——条件偏好处理效应（CPTE），并借助结构性治疗效应修饰因子这一概念，建立二者重合的条件。在第4节，我们利用匹配和分布回归等灵活构件，提出CPTE的估计策略。在第5节，我们转向基于偏好的策略学习，介绍插件法与策略价值优化方法。最后，在第6节，我们基于一步有效影响函数（EIF）校正，推导半参数有效策略估计量。我们在合成与半合成实验中验证这些方法（第7节，E 节）。我们的贡献有三方面，使基于偏好的框架内的实用策略学习成为可能。我们构建可识别的条件代理指标CPTE，用以刻画个体层面的偏好，并建立其可识别性条件。我们提出CPTE的估计程序。最后，我们基于偏好规则学习决策策略。

2 设定

因果假设。 我们假设可以获得与特征向量 $X_i \in \mathcal{X}$、处理 $T_i \in \{0,1\}$ 及观测结局$Y_i \in \mathcal{Y}$（可能是多变量的，例如$\mathcal{Y} \subset \mathbb{R}^d$）相关的 n 独立同分布样本。在潜在结果 框架下， [49], 每个个体有两个结果，$Y_i(0), Y_i(1) \in \mathcal{Y}$，但只能观测到其中一个。

我们假设 稳定单元处理值假设，无混杂性和 正值性：

假设 1（稳定单元处理值假设）。 $Y_i = Y_i(T_i)$ 对所有 $i \in [n]$。

假设 2（无混杂性）。 $\{Y_i(0),\ Y_i(1)\} \perp T_i \mid X_i$。

假设 3。 存在 $\eta \in (0,1)$，使得$\eta \leqslant e(x) \leqslant 1-\eta$ 对于所有 $x \in \boldsymbol{X}$ 均成立，其中 $e(x) = \mathbb{P}(T_i = 1 \mid X_i = x)$。

基于偏好的框架。 我们要探究一个个体在接受处理后是否会比未接受处理时情况更好。对于两个结果 $y,\ y' \in \mathcal{Y}$，“更好”并不总能用它们线性差值 $y - y'$ 的符号和大小来准确量化。因此，我们考虑一个更一般的“偏好函数”：

定义 1（偏好函数）： $w:\ (y,\ y') \in \boldsymbol{Y}^2 \mapsto w(y|y') \in \mathbb{R}$ 量化了 y 相对于y'的表现。

偏好函数的典型示例包括：（*i*）*PNS*。对于单个实值结果：

$$w(y|y') = \mathbf{1}\{y > y'\}.$$

（*ii*）一般有序结局 例如胜率比：

$$w(y|y') = \mathbb{1}_{\{y \succ y'\}} + 0.5 \cdot \mathbb{1}_{\{y \sim y'\}}, \qquad (1)$$

其中，$\succ$是 $\mathcal{Y}$上的一种序，例如 $\mathbb{R}^d$上的字典序：当且仅当 $y \succ y'$ $\inf\{k : y_k > y'_k\} < \inf\{k : y_k < y'_k\}$ 时 $y \succ y'$，用于层级结局 [44]。

（*iii*）未知偏好：如果一个“理想随机预言机”比较两个个体的结果 Y和 Y'，

$$w(Y|Y') \stackrel{\text{def}}{=} \mathbb{P}(Y \succ Y') + \tfrac{1}{2}\mathbb{P}(Y \sim Y'),$$

其中，概率反映了不确定性或个体偏好。本文假设已知 w。

3 条件效应

3.1 基于偏好的条件治疗效应的定义

个性化治疗是因果推断的终极目标。群体层面效应虽能为广泛的策略决策提供依据，却忽视了个体化策略能够处理的显著异质性。然而，由于因果推断的根本问题，个体处理效应（ITE）无法获取。在我们采用偏好函数 w的基于偏好的框架中，患者 i 的个体处理效应为：

$$\Delta_i = w(Y_i(1)|Y_i(0)).$$

它衡量 $Y_i(1)$与 $Y_i(0)$相比的情况；处理遵循 Δ_i。由于只能观测到一种行动，通常无法识别个体处理效应（ITE），

因此无法根据数据估计。个体处理效应（ITE）的常见替代指标是条件处理效应，在许多情况下可以根据数据估计。条件效应回答亚组层面的因果问题，并支持策略学习 [4], ，理想情况下能兼顾可识别性，以及针对特定人群量身定制处理方案的能力。对于风险差（RD） $w(y|y') = y - y'$，广泛使用的条件平均处理效应（CATE）恰好是个体处理效应（ITE）的条件均值:

$$\tau_{\mathrm{RD}}(x) = \mathbb{E}\left[Y_i(1) - Y_i(0) | X_i = x\right] .$$

在假设 1 至 3 下，该效应可识别，并可通过多种方法进行估计 [40, 55, 30]。遗憾的是，对于本文考虑的更一般的偏好函数，我们得到以下不可识别性结论。

引理1. 条件个体治疗效应定义如下:

$$\mathbb{E}\left[\Delta_i | X_i = x\right] = \mathbb{E}\left[w(Y_i(1)|Y_i(0)) \,|\, X_i = x\right] , \quad (2)$$

一般而言，该效应不可识别，无法从观测值的分布$(X_i, A_i, Y_i)_{i\in[n]}$中推断得出。

这是由于 w 的非可分性所致，例如 $w(y|y') \neq g(y) - g(y')$，证明见 B.1 节。当 w 可分，且可写为某个函数 g 的 $w(y|y') = g(y) - g(y')$ 时，该量就是以 $g(Y_i)$ 为结果的经典条件平均处理效应，并且可识别。请注意，引理1中的不可识别性并非源于混杂，即使在随机对照试验设置下也成立。Mueller 和 Pearl [37]、Cui 和 Han [10]、Li 和 Pearl [32]探讨了 PNS 和 QTE 的条件个体治疗效应的可识别界限。

我们提出通过定义一种可识别的条件效应，来克服不可分离条件 ITE 的不可识别性。我们将新的估计目标称为条件偏好治疗效应（CPTE）。其定义方式是对协变量进行条件化，并将反事实的联合分布 $(Y_i(1), Y_i(0))|X_i = x$ 替换为 $(Y_i(1), Y_j(0))|(X_i, X_j) = (x, x)$，此做法的灵感来自概率指数模型的形式体系 [52]。在下一小节中，我们将介绍可识别性条件；在这些条件下，CPTE 将等于条件个体治疗效应。

定义2（基于偏好的条件效**应**）。对于 $x \in \boldsymbol{X}$，令 $\{Y(0,x), Y(1,x)\}$ 为 $\{Y_i(0), Y_i(1)\}|X_i = x$ 的一个独立副本。定义

$$\begin{aligned}\mathrm{CPTE}(x) &\stackrel{def}{=} \mathbb{E}\left[w(Y(1,x)|Y_i(0)) \,|\, X_i = x\right] \\ &= \mathbb{E}\left[w(Y_i(1)|Y(0,x) \,|\, X_i = x\right] ,\end{aligned} \quad (3)$$

条件偏好处理效应。等价地，对于 $i \neq j$,

$$\mathrm{CPTE}(x) = \mathbb{E}\left[w(Y_i(1)|Y_j(0))|(X_i, X_j) = (x, x)\right] ,$$

其中，期望是针对概率分布 $Y_i(1)|X_i = x$ 和 $Y_j(0)|X_j = x$的乘积分布计算的。

在假设 1 至 3 下，

$$\mathrm{CPTE}(x) = \mathbb{E}\left[w(Y_i|Y_j) \,|\, X_i = X_j = x, T_i = 1, T_j = 0\right]$$

因此，它是可识别的。总体效应可定义为 $\mathbb{E}[$ CPTE$(X_i)]$，它是可直接折叠的（记作条件效应的总体平均值 [13, 9]）。请注意，由于独立副本总是存在，因此定义 CPTE(x)时不需要额外假设；这与 Petit 等人所定义的净获益不同[43]。最后，我们将处理效应修饰因子定义如下:

定义 3。 如果在对其他特征X^{-k}进行条件化时，变量 X^k 对应的$\mathrm{CPTE}(X_i)|X_i^k = u, X_i^{-k} = v\Big]$不是常数，则称该变量为处理效应修饰因子。不存在修饰因子时，效应是同质的；否则，效应是异质的。

处理效应修饰因子取决于效应度量 w ，并且可能不同于经典风险差的那些修饰因子。接下来，我们说明我们的CPTE与不可识别的条件个体治疗效应 $\mathbb{E}[\Delta_i|X_i = x]$之间的关系。特别地，下一小节给出了新的可识别性条件，在这些条件下 CPTE$(x) = \mathbb{E}[\Delta_i|X_i = x]$，以及用于评估当这些条件不成立时CPTE$(x)$ 与不可识别的条件个体治疗效应相差多远的工具。

3.2 CPTE(x)的性质及特征上的依赖关系

应在 X_i 中纳入哪些协变量？为使 CPTE(x) 可识别，必须满足经典因果假设（假设 1 至 3）。

尤其是，假设 2 要求必须将所有混杂因素都作为协变量纳入。治疗效应修饰因子（定义 3）是会影响治疗效应的特征。不过，它们不一定是混杂因素，也就是说，不一定是同时影响处理分配的变量。尽管如此，CPTE(x) 仍要求纳入这些因素，以捕捉总体异质性；条件效应通常也是如此。

要使公式（2）中定义的条件个体治疗效应具有可识别性，通过 等式 CPTE$(x) = \mathbb{E}[\Delta_i | X_i = x]$，仅靠处理效应修饰因子和混杂因素是不够的，正如引理1所强调的。鉴于此，我们引入结构性治疗效应修饰因子这一概念。这些变量来自结构因果模型中涵盖所有内生变量和外生变量的变量集合 Z_i，如 Pearl [42, 第 1]章所述；它们都可能使接受治疗的患者的结果优于或劣于对照患者。该定义不同于定义 3，后者仅考虑观测到的 变量。

定义 4（结构性治疗效应修饰因子）。 结构性治疗效应修饰因子是变量的一个最小集合 $I \subset [p]$，对于该集合，存在一个非恒定函数 g，使得对于所有i、 $j \in [n]$ 以及所有 $x \in \mathbb{R}^{\mathcal{I}}$:

$$\mathbb{E}\left[w(Y_i(1), Y_j(0)) \mid Z_i, Z_j, (Z_i^{\mathcal{I}}, Z_j^{\mathcal{I}}) = (x, x)\right] = g(x),$$

where $z^{\mathcal{I}} = (z_k)_{k \in \mathcal{I}}$ *for* $z \in \mathbb{R}^p$.

如下一个引理所示，如果所有结构性处理效应修饰因子都包含在协变量中，那么 CPTE(x) 与 $\mathbb{E}[w(Y_i(1)|Y_i(0))|X_i = x]$相等。然而，如果这一条件不成立，CPTE$(x)$ 可以作为 $\mathbb{E}[w(Y_i(1)|Y_i(0))|X_i = x]$的一个可识别代理指标，需要进行敏感性分析，以帮助评估我们距离相等性和可识别性 [39, 8, 23]还有多远。

定理1。 假设根据定义 *4*，结构性处理效应修饰因子包含在特征集合 X_i中。那么，对于所有 $x \in \boldsymbol{X}$, $\mathrm{CPTE}(x) = \mathbb{E}[w(Y_i(1)|Y_i(0))|X_i = x]$。

总之，为了使 CPTE 具有可识别性（在观察性研究中），特征 X_i 需要包含混杂变量、处理效应修饰因子，以捕捉治疗效应的异质性；还需要结构性治疗效应修饰因子，才能使定理1中的等式成立。然而，可能并非总能纳入甚至知晓所有结构性治疗效应修饰因子。例如，在医学背景下，这些因子可能源自遗传因素或其他未测量的变异来源。不过，正如我们在下文的实验中所展示的，即使在这种情况下，CPTE 仍可作为有用的估计目标。我们指出，另一项能够得到相同可识别性结果的假设是潜在结果独立性</style)，即对于所有 i，$Y_i(1) \perp\!\!\!\perp Y_i(0)|X_i$。这一假设认为，对照条件下的潜在结果并不能提供任何关于处理条件下潜在结果的信息，但这种假设不太可能成立。

上述定义和定理1在结构因果模型术语下的表述见附录 A.2。

3.3 示例

偏好条件 PNS。对于$\mathcal{Y} \subset \mathbb{R}$和 $w(y|y') = \mathbb{1}_{\{y>y'\}}$，我们得到一个条件且可识别的 PNS，称为偏好条件 *PNS*。这与经典 PNS [53]及其条件版本不同，后者不可识别。对协变量进行条件化，并比较独立副本中治疗组和未治疗组的结果，可得到一个可识别的估计目标，同时保留 PNS 的因果含义，并支持制定个性化策略。对于二元结局，由于伯努利分布由其均值决定，CPTE(x) 是 $\mathbb{E}[Y_i|X_i = x, A_i = a]$的直接函数，可通过标准机器学习方法进行估计。然而，对于一般的实值结果或不可识别的 PNS，不存在闭式解。在第4节中，我们将说明如何使用分布回归方法来估计这种情况下的 CPTE。

具有层级结果的条件效应。当多个结果 $\mathcal{Y}$具有层级顺序 $\succ$（例如，$\mathcal{Y} \subset \mathbb{R}^d$ 具有字典序）且 $w(y|y') = \mathbb{1}_{\{y \succ y'\}}$时，通常使用胜率比或成对比较来估计处理效应。这些方法通常针对独立 i, j 的总体层面度量 $\tau_{\text{pop}} = \mathbb{E}[w(Y_i(1)|Y_j(0))]$ [35]。此类度量用于量化治疗组个体与对照组个体之间的偏好函数，其中两者均从总体分布中随机抽取

。通过对$(X_i, X_j) = (x, x)$进行条件化，CPTE(x)便成为个性化版本，从而促进复杂结果下的策略学习。

偏好未知时的条件效应。当偏好未知时，必须学习 w 。一种常见方法是设定一个潜在奖励 R，其 $w(y|y') = \mathbb{E}[\exp(R(y) - R(y'))]$，并根据专家或参与者反馈进行估计。也可以允许 R 依赖于 X，从而得到个体特定偏好。于是，CPTE$(x) = \mathbb{E}\left[e^{R(Y_i(1))}|X_i = x\right]\mathbb{E}\left[e^{-R(Y_i(0))}|X_i = x\right]$，其中 w （以及 R）假定已学得。

4 CPTE估计方法

现在我们转向估计定义2中定义的 $x \mapsto$ CPTE(x) 。CPTE所捕捉的信息不止是反事实 $\mathbb{E}[Y_i(t)|X_i = x]$条件分布的一阶矩信息。因此，我们的估计方法会学习条件分布 $Y_i(1)|X_i$ 和 $Y_i(0)|X_i$。如果我们知道如何从$\mathcal{P}_{Y_i|T_i中X_i}$采样，那么CPTE$(x)$的一致且渐近正态估计量为：

$$\frac{1}{m}\sum_{k=1}^{m} w(\hat{y}_1^{(k)}, \hat{y}_0^{(k)}), \tag{4}$$

其中，$(\hat{y}_t^{(k)})_k$ 是从$\mathcal{P}_{Y_i|T_i=t,X_i=x}$中抽取的相互独立的随机变量，且 $(\hat{y}_1^{(k)}) \perp\!\!\!\perp (\hat{y}_0^{(k)})$。这将构成我们一般方法的基础。我们提供了几种利用现有数据 (X_i, T_i, Y_i)从$\mathcal{P}_{Y_i|T_i=t,X_i=x}$ 中抽样并估计式（4）的方法。

4.1 匹配方法

我们首先提出一种匹配方法，并证明在假设 1 至 3 及一个额外的正则性假设下，该方法具有一致性。我们采用一种 k–最近邻方法，可为CPTE(x)生成非参数相合估计量。令$\mathcal{N}_t \subset [n]$ 为接受治疗分配 $T_i = t$ 的患者集合。对于 $x \in \mathcal{X}$ 和 $k \in [n]$，令臂 $t \in \{0, 1\}$ 中 x 的 k–邻域记为 $\mathcal{N}_t(x, k)$，并将其定义为在所有大小为 k的集合$\mathcal{I} \subset \mathcal{N}_t$ 中使$\sum_{i\in\mathcal{I}}\|x - X_i\|^2$ 最小的集合。对于任意 $x \in \mathcal{X}$，令：

$$\hat{q}_W(x) \stackrel{\text{def}}{=} \frac{1}{k^2}\sum_{(i,j)\in\mathcal{N}_1(x,k)\times\mathcal{N}_0(x,k)} w(Y_i|Y_j),$$

为分布型 k–近邻估计量，满足以下条件：

定理2。假设假设 *1* 至 *3* 成立。如果函数x, $x' \mapsto \mathbb{E}[w(Y_i(1)|Y_j(0))|X_i = x, X_j = x']$ 对所有 $x \in X$均连续，则依概率有：

$$\hat{q}_W(x) \longrightarrow q_W \stackrel{def}{=} \text{CPTE}(x),$$

只要 $k \to \infty$ 和 $\frac{k}{\log_n(n)} \to 0$，其中 q_W表示 *CPTE*。

证明见 C节。对于 $x \in \mathcal{X}$ 和$(i, j) \in \mathcal{N}_1(x, k) \times \mathcal{N}_0(x, k)$，$w(Y_i|Y_j)$ 可解释为$\mathcal{P}_{Y_i|X_i=x,T_i=1} \times \mathcal{P}_{Y_j|X_j=x,T_j=0}$的一个近似样本，可用于通过公式（4）计算$\hat{}q_W$。我们也可以将'反偏好'定义为$\hat{}q_L$ ，即$w_L(y|y') = w(y'|y)$ 。与'偏好'函数 $q_W(x)$类似，$\hat{}q_L(x) \to q_L(x)$可通过分布式 K 近邻得到：

$$q_L(x) \stackrel{\text{def}}{=} \mathbb{E}[w(Y_i(0), Y_j(1))|X_i = X_j = x]. \tag{5}$$

4.2 分位数回归方法

分位数回归方法 [27]可用于参数化或非参数化地估计 $Y(1, x)$ and$Y(0, x)$ 的逆累积分布函数。这样，我们便可以从估计出的逆累积分布函数中抽取样本，用于估计 $\mathbb{E}[w(Y_i(0), Y_j(1))|X_i = X_j = x]$。在实验中，我们展示了线性分位数回归 [47, 17]和随机森林分位数回归 [36]的结果。对于任意给定的分位数模型，我们都可以使用算法 1（第 A.4 节）中描述的通用方法从逆累积分布函数中抽样，从而得到分布式 回归方法。如果 Y 是 d–维的，算法 1 还可以结合 Rosenblatt 变换 [46]，作为处理多变量结果的基础。

5 基于偏好的最优策略

5.1 策略价值

策略是一个函数 $\pi : \mathcal{X} \to \{0, 1\}$ 它将特征映射到治疗分配。对于实值结果，最优策略是在策略类 $\Pi \subset \{0, 1\}^{\mathcal{X}}$中的所有策略 π里，使策略价值 $\mathbb{E}[Y_i(\pi(X_i))]$ 最大化的策略$\pi^\star$。

如何衡量策略的偏好？我们定义**“偏好”策略价值**，用于量化策略分配的治疗相较于替代方案的可取程度。例如，临床试验可能希望估计对所有个体实施治疗相较于对照有多可取。将**“偏好”策略价值**记为，对于任意$\pi : \mathcal{X} \to \{0,\ 1\}$:

$$V(\pi) \stackrel{\text{def}}{=} \mathbb{E}\left[\pi(X_i)q_W(X_i) + (1-\pi(X_i))q_L(X_i)\right], \tag{6}$$

其中，$q_W(x) = \text{CPTE}(x)$ 和 $q_L(x)$ 与方程（5）中的相同。因此，策略学习的目标是在给定的策略类 $\Pi \subset \{0\text{中}1\}^{\mathcal{X}}$ [4]学习一个策略 π ，使“偏好”最大化:

$$\pi^\star \in \operatorname*{argmax}_{\pi\in\Pi} V(\pi)\,. \tag{7}$$

接下来，我们提出估计$\pi^\star$的不同策略。

5.2 最优治疗规则

最优治疗规则（OTR）是求解方程（7）的策略。对于无约束策略学习（即，$\Pi = \{0,\ 1\}^{\mathcal{X}}$是所有可能策略的集合），$\pi^\star$ 具有闭式表达式:

$$\forall x \in \mathcal{X},\quad \pi^\star(x) = \mathbb{1}_{\{\delta(x)>0\}}\,, \tag{8}$$

其中:

$$\forall x \in \mathcal{X},\quad \delta(x) \stackrel{\text{def}}{=} q_W(x) - q_L(x)\,. \tag{9}$$

当“偏好” $q_W(x)$ 高于“反偏好” $q_L(x)$时，最优策略会推荐一种处理。 $\delta(x)$ 也可以解释为该处理的条件净获益，对于层级结局 [6]，通常在整体人群中定义。

OTR 的直接插件估计。我们用于估计方程（8）中定义的 OTR $\pi^\star$ 的第一种策略学习方法，是直接估计方程（9）中定义的 δ 。设 $\hat{}q_W$和 $\hat{}q_L$ 为 q_W, q_L的（非参数）估计量，并令 $\hat{}\ \delta(x) \stackrel{\text{def}}{=} \hat{}\ q_W(x) - \hat{}q_L(x)$。则估计得到的 OTR 表示为:

$$\hat{\pi}_{\text{OTR}}(x) \stackrel{\text{def}}{=} \mathbb{1}_{\{\hat{\delta}(x)>0\}}\,. \tag{10}$$

根据对干扰参数 q_W 和 q_L 的估计精度，我们可以获得 $\pi^\star$的相合估计量。具体而言，如果使用 $k-$NN构建$\hat{}q_W$，$\hat{}q_L$，定理2保证$\hat{}\pi$ 是 $\pi^\star$的相合估计量。

5.3 插件估计策略价值优化

与其将δ、q_W 或 q_L 的估计量直接代入公式（8）中定义的最优治疗规则 $\pi^\star$ 的闭式表达式，不如考虑一个限制性更强的策略类 Π 进行优化。这需要估计策略的价值（定义见公式（6））。估计策略 π 的价值的一种插件估计方法是估计干扰参数$\hat{}q_W$、$\hat{}q_L$，并用定义如下的$\hat{}\ V(\pi)$估计 $V(\pi)$:

$$\hat{V}(\pi) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} \pi(X_i)\hat{q}_W(X_i) + (1-\pi(X_i))\hat{q}_L(X_i)\,. \tag{11}$$

插件估计优化方法考虑如下近似优化问题:

$$\hat{\pi}_{\text{plug-in}} \in \operatorname*{argmax}_{\pi\in\Pi} \hat{V}(\pi)\,. \tag{12}$$

可考虑如下类别 Π。

加权分类。 在我们的实验中，我们将线性分类器参数化为 f_θ ，其中 $\theta \in \Theta$，并预测 $\hat{}q_W(X_i) - \hat{}\ q_L(X_i)$ 的符号，同时按差值的绝对值加权，从而隐式地在线性分类器中最大化 $\hat{}\ V(\pi)$ 。

策略树。 也可以使用 Policytree [50] ，它直接实现方程（12），其中 Π为所有最大深度为 D（由用户指定）的二叉决策树的集合。

6 单步校正策略

对方程（11）中的干扰函数使用插件估计量，在模型具有较大的非渐近偏差（例如最近邻）、发生过拟合（例如随机森林）或模型设定错误（例如线性模型）时，可能会导致 $\hat{}\ V(\pi)$ 出现较大的偏差。有效影响函数（EIF）可用于校正这种插件偏差 [18] ，并在较弱的正则性条件下，得到具有双重稳健性和半参数效率性质的估计量。这里的效率是相对于完全非参数统计模型而言的，即基于有效影响函数的估计量在该模型中所有正则的渐近线性估计量中，具有最低的渐近方差。存在有效影响函数还要求基于偏好的策略价值参数关于观测到的

数据分布是路径可微的。与经典的策略价值估计一样，只要参数映射满足标准的光滑性条件，且偏好函数 w满足正则性要求，这一条件便成立。正值性可确保所得的有效影响函数定义良好且平方可积，并确保基于有效影响函数的估计量具有正则性，且在完全非参数模型中达到半参数效率界。因此，对于任意分布 $\mathcal{P})(o)$ 和观测值 $\mathcal{P}\ o$，我们都为偏好策略学习计算有效影响函数 $\phi($。A.3节介绍了有效影响函数校正的细节及其直观解释。

命题 1. 令 $\psi_\pi(P) = V(\pi)$ 为我们的估计量。对于 $Y(t,$ 令 $x)$ 如定义*2*中所定义：

$$p_W(x,t,y) = t\mathbb{E}\left[w(y|Y(0,x))\right] + (1-t)\mathbb{E}\left[w(Y(1,x)|y)\right],$$

类似地，将 w 替换为$w_L(y|y') = w(y'|y)$，即可得到 $p_L(x$、t和 $y)$。令 $e(x) = \mathbb{P}(T_i = 1|X_i = x)$和 q_W、q_L 如前所定义。则对于分布 P 以及 $o = (X$、T、$Y)$，ψ_π 的有效影响函数为：

$$\begin{aligned}
&\Phi_\pi(\mathcal{P})(o) = \\
&-\psi_\pi(\mathcal{P}) + \pi(X)q_W(X) + (1-\pi(X))q_L(X) \\
&\quad + \frac{T}{e(X)}\Big(\pi(X)(p_W(X,T,Y) - q_W(X)) \\
&\qquad + (1-\pi(X))(p_L(X,T,Y) - q_L(X))\Big) \\
&\quad + \frac{1-T}{1-e(X)}\Big(\pi(X)(p_W(X,T,Y) - q_W(X)) \\
&\qquad + (1-\pi(X))(p_L(X,T,Y) - q_L(X))\Big).
\end{aligned}$$

证明见 D节。令 $\hat{}e$ 为倾向评分的估计量，$\hat{}$ q_W，$\hat{}q_L$ 和 $\hat{}p_W$，$\hat{}p_L$ 为干扰参数q_W，q_L，p_W，p_L的估计量。一步策略价值估计量是$\hat{}$ $V(\pi)$（公式（6））的一阶校正，写作 $\Psi^\pi_{1-\text{step}}$ 定义为 $= \Psi^\pi_{\text{plug-in}} +_1 \sum_{i=1}^n \Phi_\pi(\hat{}\mathcal{P}_n)(O_i)$，其中 Φ_π 如命题 1所述，并将干扰参数替换为其估计量 $\hat{}e,$ $\hat{}$ q_W，$\hat{}q_L$ 和 $\hat{}p_W$，$\hat{}p_L$ （A.5节中的公式（18））。一步最优策略写作：

$$\hat{\pi}_{1-\text{step}} \in \underset{\pi\in\Pi}{\text{argmax}} \left\{\hat{V}_{1-\text{step}}(\pi) \overset{\text{def}}{=} \Psi^\pi_{1-\text{step}}\right\}. \quad (13)$$

该策略继承了一步校正策略价值的特性，预计其表现将优于插件式 OTR（5.2节）以及插件式价值优化策略（5.3节）。在实践中，策略 $\hat{}\pi_{1-\text{step}}$ 的计算方式与第5.1节相同，使用交叉拟合策略树或加权分类 $[4,\ 7]$。然而，策略价值 $\hat{}$ V_π 被其去偏价值 $\psi^\pi_{1-\text{step}}$ 所替代。在我们的实验中，与插件式价值优化策略结合使用时，一步校正始终能够提升策略表现。

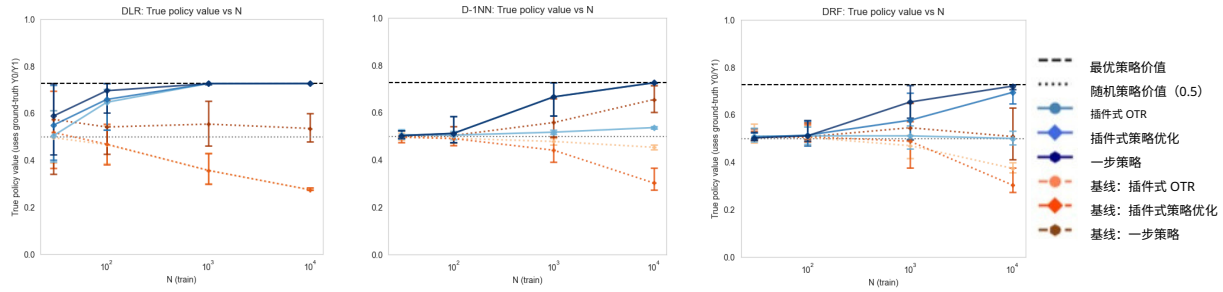
7 实验

我们通过合成和半合成实验，评估所提出方法下基于偏好的策略学习性能。该方法估计与 CPTE 相关的策略价值。结果表明，即使关键假设不成立，该方法仍能稳定地学习近最优策略。我们还将其性能与非基于偏好的方法以及具有预言机访问权限的策略进行基准比较，凸显了我们方法的稳健性和竞争力。

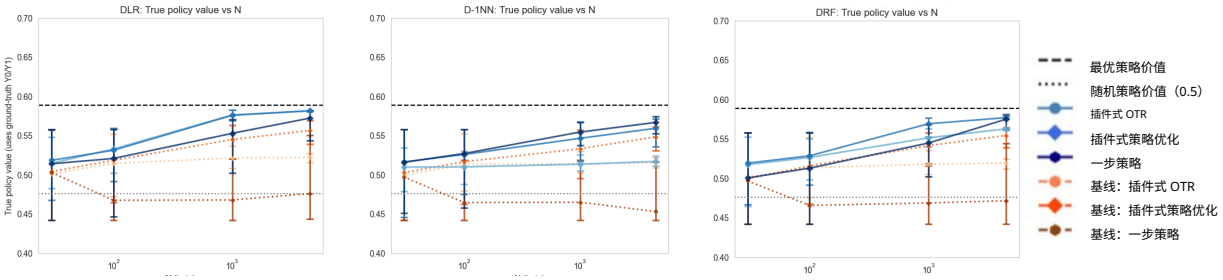
7.1 合成实验

我们考虑一个具有实值结果和 $w(y|y') = \mathbb{1}_{\{y>y'\}}$ 的合成实验设置，对应于 PNS 应用。数据生成过程详见 E.1.1 节，并满足假设 1 至 3。我们的合成实验设置探究了基于 CPTE 的偏好策略学习相较于经典 CATE 策略的优势。结果的生成方式使得每个个体的 PNS 偏好一种处理，而 CATE 则指向另一种处理。当其中一个潜在结果分布具有重尾时，就可能出现这种情况。图 2a 和 2b（E 节）展示了这类潜在结果分布的两个示例。我们的数据生成过程比较以下属性及其对基于偏好的策略学习的影响：（*i*）随机对照试验与观察性研究；（*ii*）同质与异质处理效应（定义 3）；（*iii*）反事实结果之间的相关性（负相关、不相关或正相关），这违反了定理 1中条件个体治疗效应的可识别性条件。

我们考察了不同数据生成参数组合下的策略学习与评估。我们展示了本文介绍的三种不同



(a) 合成实验。采用 PNS 策略价值进行策略学习，无相关性的异质性随机对照试验。



(b) 半合成实验。针对层级结果进行策略学习，潜在结果之间无相关性。

图 1：在结果之间不存在相关性的合成实验和半合成实验中，策略学习的结果

方法，用于从条件结果分布中抽样 $P(Y|X,\ T)$ 并估计 CPTE：如第 4.1 节所述，分布式 $k-$NN 的 k 取不同值（图 1 中$k\ =\ 1$ ）；分布随机森林（DRF）和分布式线性回归（DLR）（第 4.2 节）。我们将 CPTE 策略与使用 T 学习器框架及相应基学习器学习的 CATE 策略进行比较：标准$k-$NN、岭回归和回归森林 [30]，它们均估计 $\mathbb{E}\left[Y|X,\ T\right]$ 而非像分布方法那样从 $P(Y|X,\ T)$ 中抽样。这些方法在图中标记为基线 。随后，使用以下方法学习策略：插件式 OTR（第 5.2 节）；插件式策略价值优化（第 5.3 节）；以及使用经 EIF 校正的价值进行策略价值优化（单步校正策略，第 6 节）。策略训练 50 次，并在具有先知知识的留出集上进行评估。

在图1a中，我们展示了潜在结果不相关的异质性随机对照试验情景。该图显示了带有95%自助法置信区间的偏好策略值。EIF 校正持续改善策略学习。由于数据生成过程是线性的，DLR 表现最佳，而 DRF 和 KNN 收敛更慢。策略值估计的结果见 E.1.3 节图4。插件估计与 EIF 策略评估收敛至真实策略价值，而基于条件平均处理效应的基线策略则发散至 0 或 1。数据生成过程其他组合（随机对照试验/观察性、同质/异质、相关潜在结果）的策略学习结果见 E.1.3 节图3。重要的是，尽管相关潜在结果会显著偏倚策略值估计，学习策略本身仍能取得良好的预言机表现。即使定理 1 的可识别性假设不成立，我们的方法仍能学到良好的基于偏好的策略。

7.2 半合成实验

我们开展了一项受田纳西STAR研究 [59], 启发的半合成实验。该研究是一项具有里程碑意义的随机对照试验，旨在探究班级规模对学业成就的影响。研究追踪了田纳西州79所学校中从幼儿园到三年级的11,000多名学生，并将他们随机分配到小班、常规班或配有教师助手的常规班。研究测量了学生在包括

数学和阅读在内的核心科目成绩，并详细记录了参与度和流失情况。在我们的设定中，我们将数学成绩视为次要连续型结局，并根据学生流失情况构建二元主要结果，使用与方程（1）和 [44]中相同的偏好函数。我们将配有助教和未配助教的常规班合并，筛选后得到1,762名小班学生和4,109名常规班学生。我们通过训练多层感知机并使用MC-dropout [16], 来近似条件分布，为这些真实数据生成合成结果。这些结果随后用作最优策略和价值的真实值。完整的数据生成过程详见E.2.1节。

我们比较与合成实验中相同的估计和策略学习方法。半合成设置使我们能够控制潜在结果之间的相关性，同时仍受现实世界协变量的约束，并近似观测结果分布。我们使用 Iman-Conover 方法 [22]，对不相关的潜在结果、负相关和正相关进行了实验。半合成 STAR 数据上的结果表明，基于“偏好”的策略优于基于均值的策略，因为它们平衡了分层结果，在优先考虑流失的同时利用数学成绩。与我们的合成实验不同，偏好函数接近决策边界，使策略学习更加困难。图 1b 突出显示了这一点。DLR 被错误设定，有效影响函数在更大样本上校正了 DRF 和 KNN。基线方法表现不佳。与合成实验类似，在潜在结果之间加入强相关性不会影响策略学习性能。图 8 和图 11（E.2 节）分别汇总了所有半合成策略学习结果，并展示了相对于分层预言机策略、以及仅优化主要或次要结局的预言机策略的策略一致性结果。

8 结论

本文为策略学习以及量化最适合通过比较来评估的复杂结果条件处理效应提供了一种通用方法。我们定义*(a)* 一个不可识别的条件处理效应（方程(2)）；*(b)* 一个可识别的条件处理效应（定义2）；*(c)* 并为该可识别量提供估计量（第4节）。我们的实验表明，我们的估计量和策略学习方法能够高效地学习基于偏好的策略，从而将*(b)* 与*(c)*联系起来。它们还展示了基于偏好的策略何时可能优于基于 CATE 的策略。最后，我们通过在合成实验（第E.1.3节）和半合成实验（第E.2.2节）中为潜在结果加入相关性，探讨了违背将b 与a 联系起来的可识别性假设所产生的效应。我们发现策略评估对这一违背较为敏感，而策略学习几乎不受影响。

我们的工作为未来的若干正交研究问题铺平了道路，包括开展敏感性分析，以评估定义 2 中的条件测度与不可识别的测度接近到何种程度；以及将我们的框架用于未知的偏好函数 w，这些函数需要在优化策略的过程中进行学习。

致谢

我们感谢匿名审稿人提供的宝贵意见，这些意见帮助提升了论文的质量。M. Even 感谢 Theremia 提供的资助。D.Parnas 和 U.Shalit 得到了 ISF 项目 2456/23 的资助。

参考文献

[1] Alberto Abadie、Joshua Angrist 和 Guido Imbens。工具变量对补贴培训影响受训者收入分位数的估计。*Econometrica*，70(1)：91–117，2002。doi：10.1111/1468-0262.00270。

[2]C.M. Achilles、Helen Pate Bain、Fred Bellott、Jayne Boyd-Zaharias、Jeremy Finn、John Folger、John Johnston 和 Elizabeth Word。田纳西州学生与教师比例（STAR）项目，2008年。URLhttps://doi.org/10.7910/DVN/SIWH9F。

[3] Susan Athey 和 Guido Imbens。《异质性因果效应的递归分区》。美国国家科学院院刊，113(27):7353–7360，2016。

[4] Susan Athey 和 Stefan Wager。《基于观察性数据的策略学习》。*Econometrica*，89(1):133–161，2021。

[5]G´erard Biau 和 Luc Devroye。《最近邻方法讲义》，第 246 卷。Springer，2015。

[6] Marc Buyse。双样本问题中优先级结局的一般化成对比较。《医学统计学》，29 (30):3245–3257，2010 年 12 月。ISSN 1097-0258。doi: 10.1002/sim.3923。URL http://dx.doi.org/10.1002/sim.3923。

[7]Victor Chernozhukov、Denis Chetverikov、Mert Demirer、Esther Duflo、Christian Hansen、Whitney Newey 和 James Robins。用于处理效应和结构参数的双重／去偏机器学习，2018。

[8]B´en´edicte Colnet、Julie Josse、Ga¨el Varo- quaux 和 Erwan Scornet。目标人群中的因果效应：处理缺失协变量的敏感性分析。《因果推断期刊》，10(1):372–414，2022。

[9]B´en´edicte Colnet、Julie Josse、Ga¨el Varo- quaux 和 Erwan Scornet。风险比、优势比、风险差……哪种因果度量更容易推广? *arXiv* 预印本 *arXiv:2303.16008*，2023 年。[10] Yifan Cui 和 Sukjin Han。基于分布福利的策略学习。*arXiv* 预印本 *arXiv:2311.15878*，2023 年。[11] Johan de Aguas、Sebastian Krumscheid、Jo- han Pensar 和 Guido Biele。分层获益概率：通过稳健且稳定的推断实现部分识别。*arXiv* 预印本 *arXiv:2502.10049*，2025 年。[12] Gaohong Dong、Bo Huang、Johan Verbeeck、Ying Cui、James Song、Margaret Gamalo- Siebers、Duolao Wang、David C Hoaglin、Yo- dit Seifu、Tobias M¨ utze 等。胜率统计量（胜率比、获胜优势比和净获益）可以相互补充，展现治疗效应对事件发生时间结果的影响强度。药物统计学，22(1):20–33，2023 年。

[13] Mathieu Even 和 Julie Josse。重新审视胜率比：用于层级结果分析的因果框架。*arXiv* 预印本 *arXiv:2501.16933*，2025 年。

[14]Douglas Faries、Xiang Zhang、Zbigniew Kadziola、Uwe Siebert、Felicitas Kuehne、Robert L Obenchain 和 Josep Maria Haro。真实世界医疗数据分析：使用 *SAS* 的因果方法与实现。SAS研究所，2020。

[15] Sergio Firpo。分位数处理效应的有效半参数估计。计量经济学，75(1):259–276，2007。doi: 10.1111/j. 1468-0262.2007.00735.x。

[16] Yarin Gal 和 Zoubin Ghahramani。随机失活作为贝叶斯近似。*arXiv* 预印本 *arXiv:1506.02157*，2015。

[17]William H Greene。计量经济分析。*Pretence Hall*，2003。

[18]Oliver Hines、Oliver Dukes、Karla Diaz-Ordaz 和 Stijn Vansteelandt。揭开基于有效影响函数的统计学习之谜。《美国统计学家》，76(3): 292–304，2022。

[19] Paul W. Holland。《统计学与因果推断》。美国统计学会杂志，81(396):945–960，1986。ISSN 01621459，1537274X。URL http://www.jstor.org/stable/2289064。[20] Chia-Ling Huang、John Lamb、Leonid Chindelevitch、Jarek Kostrowicki、Justin Guinney、Charles DeLisi 和 Daniel Ziemek。相关性集合分析：利用先验因果知识检测疾病人群中的活跃调控因子。*BMC*生物信息学，13(1):46，2012。[21] Emily J Huang、Ethan X Fang、Daniel F Hanley 和 Michael Rosenblum。治疗获益的不平等：我们能否确定一种新处理是使多数人还是少数人获益？生物统计学，18(2):308–324，2017。[22] Ronald L Iman 和 William-Jay Conover。诱导秩的无分布方法

输入变量之间的相关性。《统计学—模拟与计算》，11(3):311–334，1982。

[23]Ying Jin、Zhimei Ren 和 Emmanuel J Cand`es。个体处理效应的敏感性分析：一种稳健的保形推断方法。《美国国家科学院院刊》，120(6):e2214889120，2023 年。

[24]Nathan Kallus 和 Miruna Oprescu。条件分布处理效应的稳健且无偏学习。载于人工智能与统计国际会议，第 6037–6060 页。PMLR，2023 年。

[25]Yuta Kawakami、Manabu Kuroki 和 Jin Tian。连续变量和向量变量的因果概率。*arXiv* 预印本 *arXiv:2405.20487*，2024 年。

[26]Christian Kleiber 和 Achim Zeileis。使用 *R* 进行应用计量经济学。Springer-Verlag，纽约，2008 年。doi：10.1007/978-0-387-77318-6。URL `https://CRAN.R-project.org/package=AER`。

[27]Roger Koenker 和 Gilbert Bassett。回归分位数。*Econometrica*，46(1):33，1978年1月。ISSN 0012-9682。doi: 10. 2307/1913643。URL `http://dx.doi.org/10.2307/1913643`。

[28] Roger Koenker 和 Gilbert Bassett。回归分位数。*Econometrica*，46(1):33–50，1978年。doi: 10.2307/1913643。

[29] Alan B Krueger。教育生产函数的实验估计。经济学季刊，114(2):497–532，1999年。

[30] S¨oren R K¨unzel、Jasjeet S Sekhon、Peter J Bickel 和 Bin Yu。利用机器学习估计异质性处理效应的元学习器。《美国国家科学院院刊》，116(10):4156– 4165，2019年。

[31]Lihua Lei. 秩回归的因果解释。*arXiv* 预印本 *arXiv:2406.05548*，2024。

[32] Ang Li 和 Judea Pearl。非二元处理和效应下的因果概率。载于人工智能促进会会议论文集，第 38 卷，第 20465–20472 页，2024 年。[33] Jiannan Lu、Peng Ding 和 Tirthankar Dasgupta。序数结果的处理效应：因果估计目标与紧界。《教育与行为统计学杂志》，43（5）：540–567，2018 年。[34] Charles F Manski。概率分布的部分识别。Springer，2003 年。

[35]Lu Mao。关于使用 u- 统计量进行因果估计。*Biometrika*，105(1):215–220，2017 年 12 月。ISSN1464-3510。

[36]Nicolai Meinshausen 和 Greg Ridgeway。分位数回归森林。机器学习研究期刊，7(6)，2006。

[37] ScottMuellerandJudeaPearl。个性化决策——概念性介绍。因果推断期刊，11(1):20220050，2023。

[38] Jeffrey N¨af 和 Herbert Susmann。Causal-drf: 使用分布式随机森林进行条件核处理效应估计。*arXiv* 预印本 *arXiv:2411.08778*，2024。

[39] Trang Quynh Nguyen、Benjamin Ackerman、Ian Schmid、Stephen R Cole 和 Elizabeth A Stuart。当目标总体中未观测到效应修饰因子时，对随机对照试验中治疗效应泛化进行敏感性分析：假设、模型、效应尺度、数据情景及实施细节。*PLOS ONE*，13(12): e0208795，2018。

[40] Xinkun Nie 和 Stefan Wager。异质性处理效应的准神谕估计。*Biometrika*，108(2)：299–319，2021。

[41]Junhyung Park、Uri Shalit、Bernhard Sch¨olkopf 和 Krikamol Muandet。基于核条件均值嵌入和 U统计量回归的条件分布治疗效应。载于 Marina Meila 和 Tong Zhang（编），第 *38* 届

国际机器学习会议，第 139 卷，机器学习研究论文集，第 8401–8412 页。PMLR，2021 年 7 月 18–24 日。URL https://proceedings.mlr.press/v139/park21c.html。

[42] Judea Pearl。《因果关系》。剑桥大学出版社，英格兰剑桥，2009年9月。

[43]Fran¸cois Petit、G´erard Biau 和 Rapha¨el Porcher。实现处理净获益最大化的最优治疗方案。*arXiv* 预印本 *arXiv:2503.22580*，2025年。

[44]S. J. Pocock、C. A. Ariti、T. J. Collier 和 D. Wang。胜率比：一种基于临床优先事项分析临床试验复合终点的新方法。《欧洲心脏杂志》，33(2):176–182，2011年9月。ISSN 1522-9645。doi: 10.1093/eurheartj/ehr352。URL http://dx.doi.org/10.1093/eurheartj/ehr352。

[45] James M Robins 和 Thomas S Richardson。替代图形因果模型与直接效应的识别。《因果关系与精神病理学：寻找疾病及其治疗方法的决定因素》，84:103–158，2010。[46] Murray Rosenblatt。关于多元变换的评论。《数学统计年鉴》，23(3):470–472，1952。[47] Skipper Seabold 和 Josef Perktold。statsmodels：使用 Python 进行计量经济学和统计建模。载于第 *9* 届 *Python*科学会议，2010。[48] Ryusei Shingaki 等。利用协变量信息识别和估计观察性研究中潜在结果的联合概率。《神经信息处理系统进展》，34:26475–26486，2021。[49] Jerzy Splawa-Neyman、Dorota M Dabrowska 和 Terrence P Speed。概率论在农业实验中的应用：关于原则的论文，第 9 节。《统计科学》，第 465–472 页，1990。[50]Erik Sverdrup、Ayush Kanodia、Zhengyuan Zhou、Susan Athey 和 Stefan Wager。policytree：通过双重稳健经验福利最大化进行策略学习。《开源软件期刊》，5(50):2232，2020。

[51] Xiaoqing Tan、Shu Yang、Wenyu Ye、Douglas E Faries、Ilya Lipkovich 和 Zbigniew Kadziola。用于估计平均处理效应的双重机器学习方法：一项比较研究。2025 年。

[52] Olivier Thas、Jan De Neve、Lieven Clement 和 Jean-Pierre Ottoy。概率指数模型。英国皇家统计学会会刊 *B* 辑：统计方法学，74(4)：623–671，2012 年。

[53] Jin Tian 和 Judea Pearl。因果概率：界限与识别。数学与人工智能年刊，28 (1)：287–313，2000 年。

[54] Aad W Van der Vaart。渐近统计学，第 3 卷。剑桥大学出版社，2000 年。

[55] Stefan Wager 和 Susan Athey。使用随机森林估计异质性处理效应并进行推断。《美国统计协会杂志》，113(523): 1228–1242，2018。[56] Stefan Wager、Susan Athey 和 Ryan Tibshirani。广义随机森林。统计年刊，47(2):1148–1178，2019。[57] Shuai Wang 和 Ang Li。使用机器学习模型估计因果概率。*arXiv* 预印本 *arXiv:2502.08858*，2025。[58] Shuai Wang、Song Jiang、Yizhou Sun、Judea Pearl 和 Ang Li。从有限总体数据中学习因果概率。*arXiv* 预印本 *arXiv:2505.17133*，2025。[59]Elizabeth Word、John Johnston、Helen P Bain、B DeWayne Fulton、Jayne B Zaharias、Charles M Achilles、Martha N Lintz、John Folger 和 Carolyn Breda。田纳西州学生/教师成就比率（STAR）项目现状。田纳西州教育委员会，1990。[60]Tianhui Zhou、William E Carson IV 和 David Carlson。估计潜在结果

分布与协作式因果网络。机器学习研究汇刊，
2022： https–openreview，2022。

A 进一步的具体细节

A.1 说明性对抗示例

示例 1：观测到的和结构性处理效应修饰因子。 设想一个简单的设定，其中我们假定特征 $X_i \in \mathcal{X}$ 不 包含处理效应修饰因子，记为 $O_i \in \mathcal{O}$。假设 $\mathcal{Y} \subset \mathbb{R}$, $w(y|y') = \mathbb{1}_{\{y>y'\}}$，处理效应修饰因子集合 $\mathcal{O}$ 具有以下形式：$\mathcal{O} = \mathcal{O}_1 \cup \mathcal{O}_2$，其中 $\mathbb{P}(O_i \in \mathcal{O}_1) = 1-\alpha$, $\mathbb{P}(O_i \in \mathcal{O}_2) = \alpha$，且 O_i 与 X_i相互独立，使得：

$$Y_i(1) = y_1 > y_0 = Y_i(0) | O_i \in \mathcal{O}_1 \,,$$
$$Y_i(1) = y_1' < y_0' = Y_i(0) | O_i \in \mathcal{O}_2 \,,$$

几乎必然成立。因此，如果 $y_0' > y_1 > y_0 > y_1'$，我们有：

$$\mathbb{E}\left[w(Y(1, X_i)|Y_i(0))|X_i\right] = (1-\alpha)^2 \,, \tag{14}$$
$$\mathbb{E}\left[w(Y_i(1)|Y_i(0))|X_i\right] = 1-\alpha \,, \tag{15}$$
$$\mathbb{E}\left[w(Y(1, X_i)|Y_i(0))|X_i, O_i\right] = \mathbb{1}_{\{O_i \in \mathcal{O}_1\}} \,. \tag{16}$$

方程 (14) 是仅观测到 X_i 时的 CPTE，而它不提供信息。如果我们还观测到信息性变量 O_i，则在方程 (16) 中，异质性由处理效应修饰因子 O_i 完全捕获。然而，方程 (15) 也可以写成在给定 X_i 条件下方程 (16) 的均值：但得到的价值与方程 (14) 中的不同。这是因为 O_i 也是一个结构性 处理效应修饰因子。

示例 2：线性条件平均治疗效应 与 CPTE。现在我们通过一个示例，展示经典线性条件平均治疗效应与我们提出的基于偏好的条件效应在治疗建议上的差异。该示例说明了 CPTE 如何利用一阶矩信息以外的信息，并在提出治疗建议时采用多数投票方法。设 $\mathcal{Y} = \mathbb{R}$ 和

$$Y_i(1) \sim \frac{\delta_0 + \delta_{0.1} + \delta_{1.1}}{3} | X_i = x \,,$$
$$Y_i(0) \sim \frac{\delta_{-0.1} + \delta_1}{2} | X_i = x \,.$$

在这种情况下，CPTE$(x) = 0.66 > 0.5$，而 $\mathbb{E}\left[Y_i(1) - Y_i(0)|x\right] = 0.4 - 0.45 = -0.05 < 0$。治疗建议会因决策者的目标而异——他们是考虑效应的大小（风险差），还是考虑最可能使个体 X_i获益的处理。本原型构成了我们合成实验的基础。

A.2 转换为结构因果模型（SCM）术语。

这里，我们在 Pearl 的结构因果模型（SCM）框架内重新表述若干基本定义和定理。设 $(\mathcal{G}\ U,\ f)$ 为一个结构因果模型，其中 U 表示外生变量，$V = (T,\ Y,\ X)$ 表示内生变量，且

$$Y = f_Y(T, X, U_Y).$$

反事实结果由干预定义为

$$Y(t) := Y \mid \mathrm{do}(T = t).$$

基于偏好的个体因果效应。 对于由外生变量 U 刻画的个体，个体层面的基于偏好的因果效应为

$$\Delta(U_i) := w\big(f_Y(1, X_i, U_i),\ f_Y(0, X, U_i)\big) \,.$$

注意，我们确实有 $\Delta_i = \Delta(U_i)$。因此，条件个体处理效应等于 :

$$\mathbb{E}\left[w\big(f_Y(1, X_i, U_i),\ f_Y(0, X, U_i)\big)\right] \,.$$

CPTE 的结构因果模型解释 - 定义2。 对于给定的协变量取值 $X = x$，条件偏好处理效应 CPTE(x) 对应于比较在外生变量独立抽样下生成的结果：

$$U \sim P(U \mid X = x), \qquad U' \sim P(U \mid X = x), \qquad U \perp U',$$

以及定义2中定义的 $Y(1,x), Y(0,x)$ 也可以定义为：

$$Y(1,x) = f_Y(1, x, U_Y), \qquad Y(0,x) = f_Y(0, x, U'_Y),$$

从而得到相同的 CPTE 定义：

$$\mathrm{CPTE}(x) = \mathbb{E}\big[w\big(Y(1,x), Y_i(0)\big)\big] = \mathbb{E}\big[w\big(Y_i(1), Y(0,x)\big)\big]$$

结构性治疗效应修饰因子（结构因果模型）- 定义 4。 一组变量 $Z_I \subseteq U$ 是一组结构性治疗效应修饰因子，如果对于所有 $i,\ j \in [n]$ 以及所有 $x \in \mathbb{R}^{\mathcal{I}}$：

$$\mathbb{E}\left[w(Y_i(1), Y_j(0)) \mid Z_i, Z_j, (Z_i^{\mathcal{I}}, Z_j^{\mathcal{I}}) = (x, x)\right] = g(x)\,,$$

其中 $z^{\mathcal{I}} = (z_k)_{k \in \mathcal{I}}$ 对于 $z \in \mathbb{R}^p$。

CPTE 等价的条件　　q于个体层面估计——定理1。
假设按照定义 4 的含义，结构性治疗效应修饰因子包含在特征集合 X_i 中。那么，对于所有 $x \in \mathcal{X}$，CPTE$(x) = \mathbb{E}\left[w(Y(1,U_i)|Y(0,U_i))|X_i = x\right]$。

A.3 使用有效影响函数校正插件偏差

方程（11）中干扰函数的插件估计量可能导致 $\hat{}\ V(\pi)$ 出现较大偏差 [18]，如果模型存在较大的非渐近偏差（例如最近邻），如果模型过拟合（例如随机森林），或者模型设定错误（例如线性模型）。有效影响函数（EIF）可用于校正这种插件偏差 [18]，并获得有效估计量；在温和的假设下，这些估计量具有双重稳健性、半参数效率最优性，并且在某些估计量类别中具有最小渐近方差。

设 $\Psi(\mathcal{P})$ 是我们希望估计的估计目标，也就是观测分布$\mathcal{P}$的函数。在我们的情形中，$\mathcal{P}$ 是观测值 n $O_i = (X_i, T_i, Y_i)$的分布，而估计目标为$\Psi_\pi(\mathcal{P}) = V(\pi)$，因此：

$$\Psi_\pi(\mathcal{P}) = \mathbb{E}\left[\pi(X_i)q_W(X_i) + (1 - \pi(X_i)q_L(X_i))\right].$$

如前所述，为了估计 $\Psi_\pi(\mathcal{P})$，我们首先需要在数据上学习得到（非参数）干扰项估计（即 q_W和 q_L的估计）。然而，这些干扰项估计中的微小扰动和误差，如果我们只是像公式（6）那样直接代入干扰项估计量，就可能导致对 Ψ_π 的估计产生偏差。这就是 Ψ_π的插件策略估计量，记为 $\Psi^\pi_{\mathrm{plug-in}}$。

有效影响函数（EIF）。 Ψ_π 在$\mathcal{D}$ 处、对任意观测 o的有效影响函数 $\phi_\pi(\mathcal{P})$ 可以写为：

$$\phi(\mathcal{P})(o) \stackrel{\mathrm{def}}{=} \lim_{t \to 0} \frac{\mathrm{d}\psi(\mathcal{P}_t)}{\mathrm{d}t}\,, \quad \text{where} \quad \mathcal{P}_t = (1-t)\mathcal{P} + t\delta_o\,,$$

其中 δ_o 是位于 o的狄拉克。应将其理解为 Ψ_π的“梯度”，在$\mathcal{P}$ 处计算，并在观测 o上求值。Hines 等人 [18] 提供了计算这些有效影响函数的方法，并以平均处理效应（ATE）、平均处理组效应（ATT）或风险比（RR）等经典目标估计量为例。一步估计量校正了因错误估计分布的干扰参数而产生的一阶偏差，其中 $\Psi^\pi_{\mathrm{plug-in}} = \hat{}V(\pi)$（方程（11））， 且 Φ^π 是其有效影响函数：

$$\Psi^\pi_{1-\mathrm{step}} \stackrel{\mathrm{def}}{=} \Psi^\pi_{\mathrm{plug-in}} + \frac{1}{n}\sum_{i=1}^{n} \Phi^\pi(\hat{\mathcal{P}}_n)(O_i)\,. \tag{17}$$

A.4 分布回归算法

元算法算法 1根据给定的分位数模型 $M_q^{(t)}(x)$ 和插值规则，输出 $q_W(x)$ $=\mathbb{E}\left[w(Y_i(1), Y(0, X_i))|X_i = x\right] = \mathrm{CPTE}(x)$ 以及 $q_L(x) = \mathbb{E}\left[w(Y_i(0), Y(1, X_i))|X_i = x\right]$（治疗反转后的CPTE）的估计量 $\hat{}q_W$、$\hat{}q_L$：$\mathrm{X} \mapsto$ $[0,1]$。首先，算法 1将 $[0,1]$ 离散化为Q，并计算与这些分位数对应的反事实的价值 $Y_q(t,x)$ 。该步骤估计逆累积分布函数。然后对 S 个其他分位数进行抽样，并使用插值规则、从逆累积分布函数抽样，计算与这些分位数对应的价值。 `Interp` 是任意函数，以集合 $(q, f(q))_{q\in\mathcal{Q}}$ 和价值 $x \in [0,1]$ 为输入，其中Q是 $[0,1]$, f 的离散化： $[0,1] \to$ Y是仅在离散化价值上求值的模型。 `Interp` 随后为 $f(x)$输出一个价值。最后一步对所得价值取平均。

算法 1 通过分位数模型对 q_W， q_L 进行基于抽样的估计

要求： 分位数模型 $M_q^{(1)}(x)$ 和 $M_q^{(0)}(x)$; 网格大小 $|\mathcal{Q}|$; 样本数 S; 插值规则 `Interp`。

确保： 估计量 q_W， q_L。

1: 生成一个网格 $\mathcal{Q} = \{q_1, \ldots, q_{|\mathcal{Q}|}\}$，其中包含 $q_k \sim$ 个 Unif$(0,\ 1)$。2: 对于每个 $q \in \mathcal{Q}$，计算

$$Y_q(1,x) \leftarrow M_q^{(1)}(x), \qquad Y_q(0,x) \leftarrow M_q^{(0)}(x).$$

3: 抽取 S 个独立分位数 $u_s^{(1)}$， $u_s^{(0)} \sim$ 个 Unif$(0,\ 1)$，其中 $s = 1, \ldots,$ S。4: 对于每个 s，进行插值：

$$Y_s(1,x) \leftarrow \texttt{Interp}\big(\{(q, Y_q(1,x))\}_{q\in\mathcal{Q}},\, u_s^{(1)}\big),$$

$$Y_s(0,x) \leftarrow \texttt{Interp}\big(\{(q, Y_q(0,x))\}_{q\in\mathcal{Q}},\, u_s^{(0)}\big).$$

5: 估计

$$\widehat{q}_W(x) \leftarrow \frac{1}{S}\sum_{s=1}^{S} w\big(Y_s(1,x),\, Y_s(0,x)\big).$$

$$\widehat{q}_L(x) \leftarrow \frac{1}{S}\sum_{s=1}^{S} w\big(Y_s(0,x),\, Y_s(1,x)\big).$$

A.5 一步策略价值估计量

式（17）中定义的一步策略价值估计量写作：

$$\begin{aligned}\Psi^{\pi}_{1-\mathrm{step}} = \frac{1}{n}\sum_{i=1}^{n}\Big[&\pi(X_i)\hat{q}_W(X_i) + (1-\pi(X_i))\hat{q}_L(X_i)\\ &+ \frac{T_i}{e(X_i)}\Big(\pi(X_i)(\hat{p}_W(X_i,T_i,Y_i) - \hat{q}_W(X_i))\\ &+ (1-\pi(X_i))(\hat{p}_L(X_i,T_i,Y_i) - \hat{q}_L(X_i))\Big)\\ &+ \frac{1-T_i}{1-e(X_i)}\Big(\pi(X_i)(\hat{p}_W(X_i,T_i,Y_i) - \hat{q}_W(X_i))\\ &+ (1-\pi(X_i))(\hat{p}_L(X_i,T_i,Y_i) - \hat{q}_L(X_i))\Big)\Big].\end{aligned} \tag{18}$$

B 其他证明

B.1 引理1的证明

证明。 假设$\mathcal{Y} = \{0,1\}$，使得 $Y_i(1), Y_i(0)$ 是均值为 0.5 的伯努利变量。我们取$Y_i(1), Y_i(0)$，使其独立于 X_i。将 $(Y_i(1), Y_i(0))|X_i = x$ 耦合为 $Y_i(1) = Y_i(0)$，可得 $\mathbb{P}(Y_i(1) > Y_i(0)) = 0$；而取相互独立的潜在结果则得到 $\mathbb{P}(Y_i(1) > Y_i(0)) = \mathbb{P}(Y_i(1) = 1, Y_i(0) = 0) = 1/4$。由于无论采用哪种耦合，观测值的分布 (X_i, T_i, Y_i) 均不变，但 $\mathbb{E}[w(Y_i(1)|Y_i(0))]$ 的价值会变化，因此存在不可识别性：
$\mathbb{E}[w(Y_i(1)|Y_i(0))]$ 无法表示为观测数据的分布的函数。 □

B.2 定理1的证明

证明。 如果 $X_i = Z_i^{\mathcal{I}}$，其中 $\mathcal{I}$ 为结构性治疗效应修饰变量集合，或者如果 $Z_i \subset X_i$，则存在 g 使得 $\mathbb{E}[w(Y_i(1)|Y_j(0))|Z_i,\ Z_j,\ (X_i,\ X_j) = x] = g(x)$。于是，

$$\begin{aligned}\mathbb{E}[w(Y_i(1)|Y_j(0)|X_i = X_j = x] &= \mathbb{E}\left[\mathbb{E}\left[w(Y_i(1)|Y_j(0)|Z_i, Z_j, (Z_i^{\mathcal{I}}, Z_j^{\mathcal{I}}) = (x,x)\right]|X_i = X_j = x\right] \\ &= \mathbb{E}[g(x)|X_i = X_j = x] \\ &= g(x)\,.\end{aligned}$$

现在，取 $i = j$,

$$\begin{aligned}\mathbb{E}[w(Y_i(1)|Y_i(0)|X_i = x] &= \mathbb{E}\left[\mathbb{E}\left[w(Y_i(1)|Y_i(0)|Z_i, Z_i^{\mathcal{I}} = x\right]|X_i = x\right] \\ &= \mathbb{E}[g(x)|X_i = x] \\ &= g(x)\,,\end{aligned}$$

证明得证。 □

C 定理2的证明

设$\mathbf{X_t} = \{X_i, T_i = t\}$ 对应于处理组 $t = 0, 1$以及$\mathbf{X} = \mathbf{X}_0 \cup \mathbf{X}_1$。对于任意 $x \in \mathcal{X}$，令 $Y(1,x)$ 与 $Y(0,x)$ 分别为 $Y_i(1)|X_i = x$ 和 $Y_i(0)|X_i = x$ 的独立副本，且它们彼此独立。 $Y(1,x)$ 与 $Y(0,x)$ 相互独立。

设某点 $x \in \mathcal{X}$ 的处理组 $t \in \{0,1\}$ 的 k–最近邻域为：

$$\mathcal{N}_t(x,k) = \{a \in \mathbf{X_t}\ s.t.\ |\{b \in \mathbf{X_t} : ||b - x|| < ||a - z||\}| < k\}$$

一个一般的邻 域，独立于治疗分配，记为$\mathcal{N}(x,k)$。首先，

$$\begin{aligned}\mathbb{E}[\hat{q}(x)] &= \mathbb{E}[w(Y_i|Y_j)|(i,j) \in \mathcal{N}_1(x,k) \times \mathcal{N}_0(x,k)] \\ &= \mathbb{E}[w(Y_i(1)|Y_j(0))|(i,j) \in \mathcal{N}_1(x,k) \times \mathcal{N}_0(x,k)] \quad \text{(Assumptions 1 and 2)}\,.\end{aligned}$$

步骤1： $\mathbb{E}[\hat{}q(x)] \to q(x)$。我们有：

$$\begin{aligned}\mathbb{E}[\hat{q}(x)] - q(x) &= \mathbb{E}[w(Y_i(1)|Y_j(0))|(i,j) \in \mathcal{N}_1(x,k) \times \mathcal{N}_0(x,k)] - \mathbb{E}[w(Y(1,x), Y(0,x))] \\ &= \mathbb{E}[w(Y_i(1)|Y_j(0)) - w(Y(1,x), Y(0,x))|(i,j) \in \mathcal{N}_1(x,k) \times \mathcal{N}_0(x,k)]\,.\end{aligned}$$

We 然后介绍以下经典引理，并为完整起见给出证明 .

引理 2. 设 $x \in X$ 和任意 $\hat{}x \in N(x,k)$。我们有：

$$\hat{x} \xrightarrow{p} x$$

当 $k = o(n/log(n))$ 时。

证明。我们需要证明:

$$\forall \epsilon > 0\,, \quad \lim_{n\to\infty} P(||x - \hat{x}|| > \epsilon)\,.$$

令 $B(x,\epsilon)$ 为以 $\mathbf{x}$ 为中心、半径为 ϵ 的球体，并令 $p = \mathbb{P}(\hat{}x \in B(x,\epsilon))$ 为 $\hat{}x$ 落在 ϵ-球体内的概率。我们将落在 ϵ-球体内的数据点数量记为 $Z = |\{z \in \mathbf{X} | z \in B(x,\epsilon)\}|$。注意到 $Z \sim Bin(N,p)$，因此:

$$\lim_{n\to\infty} P(||x - \hat{x}|| > \epsilon) \leqslant \lim_{n\to\infty} \mathbb{P}(Z < k) = \lim_{n\to\infty} \sum_{i=0}^{k-1} \binom{n}{i} p^i (1-p)^{n-i}$$

如果 $p \in (0,1)$,

$$\lim_{n\to\infty} \sum_{i=0}^{k-1} \binom{n}{i} p^i (1-p)^{n-i} \leqslant \lim_{n\to\infty} k n^k q^n$$

其中 $q = \mathsf{max}(p, 1-p) < 1$，并且我们将 $\binom{n}{i}$ 的上界设为 $n^i \leqslant n^k$。因此，由于 $kn^k q^n = \mathsf{exp}(\mathsf{log}(k) + l\,\mathsf{log}(n) + n\,\mathsf{log}(q))$ 且 $\mathsf{log}(q) < 0$，只要 $\frac{k}{\log_n(n)} \to 0$:

$$\lim_{n\to\infty} P(||x - \hat{x}|| > \epsilon)\,,$$

由此得证。接下来，若 $p \notin (0,1)$，则 $p = 0$ 或 $p = 1$。$p = 0$ 对应于 $x \notin \mathsf{Supp}(\mathrm{X_i})$，这是我们不予考虑的情形（因为 q 将在抽样的数据点上进行评估）；而若 $p = 1$，则意味着 X_i 是一个等于 x 的确定性随机变量，也就是说该引理显然成立。 □

由于我们假设 $w: (x,x') \to q(x,x')$ 几乎处处连续，因此根据连续映射定理 [54]，我们得出 $\mathbb{E}[\hat{} q(x)] \to q(x)$。为保证一致性，注意到

$$\mathbb{P}\left(|\hat{q}(x) - q(x)| > \epsilon\right) = \mathbb{P}\left(|\hat{q}(x) - \mathbb{E}[\hat{q}(x)] + \mathbb{E}[\hat{q}(x)] - q(x)| > \epsilon\right)\,.$$

利用并集界,

$$\mathbb{P}\left(|\hat{q}(x) - q(x)| > \epsilon\right) \leqslant \mathbb{P}\left(|\hat{q}(x) - \mathbb{E}[\hat{q}(x)]| > \epsilon/2\right) + \mathbb{P}\left(|\mathbb{E}[\hat{q}(x)] - q(x)| > \epsilon/2\right)\,.$$

因此，我们需要证明:

$$\forall \varepsilon > 0\,, \quad \mathbb{P}\left(|\hat{q}(x) - \mathbb{E}[\hat{q}(x)]| > \epsilon\right) \to 0\,.$$

步骤 2: $|\hat{} q(x) - \mathbb{E}[\hat{} q(x)]| \to 0$ **依概率收敛**。我们有:

$$\begin{aligned}\mathbb{P}\left(|\hat{q}(x) - \mathbb{E}[\hat{q}(x)]| > \epsilon\right) &= \mathbb{P}\left(\left|\frac{1}{k^2} \sum_{(i,j)\in\mathcal{N}_1(x,k)\times\mathcal{N}_0(x,k)} w(Y^i, Y^j) - \mathbb{E}[w(Y^i, Y^j)]\right| > \epsilon\right) \\ &\leqslant \frac{\mathbb{E}\left[\left(\sum_{(i,j)\in\mathcal{N}_1(x,k)\times\mathcal{N}_0(x,k)} w(Y^i, Y^j) - \mathbb{E}[w(Y^i, Y^j)]\right)^2\right]}{k^4\epsilon^2}\,,\end{aligned}$$

利用切比雪夫不等式。我们记 $d(i,j) := w(Y^i, Y^j) - \mathbb{E}[w(Y^i, Y^j)]$。我们可以将前面的表达式改写为:

$$\frac{\mathbb{E}\left[\left(\sum_{(i,j)\in\mathcal{N}_1(x,k)\times\mathcal{N}_0(x,k)} d(i,j)\right)^2\right]}{k^4\epsilon^2}\,.$$

我们根据偏好函数中进行比较的配对之间是否存在重叠性˜，将二次表达式分为两组。我们将四重交叉乘积记为$\mathcal{N}^{} = \mathcal{N}_1(x,\ k) \times \mathcal{N}_1(x,\ k) \times \mathcal{N}_0(x,\ k) \times \mathcal{N}_0(x,\ k)$。

$$\frac{1}{k^4\epsilon^2} \sum_{(i,\hat{i},j,\hat{j})\in\tilde{N}} \left(\mathbb{I}\{i = \hat{i}\ or\ j = \hat{j}\} + \mathbb{I}\{i \neq \hat{i}\}\mathbb{I}\{j \neq \hat{j}\}\right) \mathbb{E}[d(i,j)d(\hat{i},\hat{j})]$$

情形 1</style}：存在 $2k^3$ 种情形，其中至少有 $i = \hat{i}$ 或 $j = \hat{j}$。

$$\begin{aligned}
&\mathbb{E}[d(i,j)d(\hat{i},\hat{j})] \\
&= \operatorname{cov}[w(Y^i, Y^j), w(Y^{\hat{i}}, Y^{\hat{j}})] \qquad \text{(definition of covariance)} \\
&\leqslant \operatorname{var}[w(Y^i, Y^j)] \\
&\leqslant 1\,,
\end{aligned}$$

因为假设 w 有界于 1。因此，该求和可以界定如下：

$$\begin{aligned}
\frac{1}{k^4\epsilon^2} \sum_{(i,\hat{i},j,\hat{j})\in\tilde{N}} \mathbb{I}\{i = \hat{i} \text{ or } j = \hat{j}\}\mathbb{E}[d(i,j), d(\hat{i},\hat{j})] &\leqslant \frac{2k^3}{k^4\epsilon^2} \\
&= \frac{2}{k\epsilon^2}\,.
\end{aligned}$$

情形 2：当 $i \neq \hat{i}$ 且 $j \neq \hat{j}$时，共有 $O(k^4)$ 种情形。由于这些配对是独立同分布的，若没有共同成员，协方差为零。因此，

$$\begin{aligned}
\frac{1}{k^4} \sum_{(i,\hat{i},j,\hat{j})\in\tilde{N}} \mathbb{I}\{i \neq \hat{i}\}\mathbb{I}\{j \neq \hat{j}\}\mathbb{E}[d(i,j)d(\hat{i},\hat{j})] &= \frac{1}{k^4} \sum_{(i,\hat{i},j,\hat{j})\in\tilde{N}} \mathbb{I}\{i \neq \hat{i}\}\mathbb{I}\{j \neq \hat{j}\}\, Cov[w(Y^i, Y^j), w(Y^{\hat{i}}, Y^{\hat{j}})] \\
&= 0\,.
\end{aligned}$$

因此，

$$\mathbb{P}\left(|\hat{q}(x) - \mathbb{E}[\hat{q}(x)]| > \epsilon\right) \leqslant \frac{2}{k\epsilon^2}\,,$$

只要 $k \to \infty$，对于任意固定的 $\varepsilon > 0$，该式就收敛于 0。证明至此完成。

D 命题 1 的证明

命题 2。令 $\psi_\pi(P) = V(\pi)$ 为我们的估计目标。

1. 对于 $\pi(x) = 1$ 常数策略，$\psi_\pi = \psi_W$ 是获胜比例，即主要干预优于替代方案的可能性。其有效影响函数可写为；对于 $o = (X、\ T、\ Y)$:

$$\begin{aligned}
\phi_W(\mathcal{P})(o) = &-\psi_W(\mathcal{P}) + \mathbb{E}\left[w(Y(1,X_i)|Y(0,X_i))|X_i = X\right] \\
&+ \frac{T}{e(X)}\left(\mathbb{E}\left[w(Y|Y(0,X_i))\right] - \mathbb{E}\left[w(Y(1,X_i)|Y(0,X_i))|X_i = X\right]\right) \\
&+ \frac{1-T}{1-e(X)}\left(\mathbb{E}\left[w(Y(1,X_i)|Y)\right] - \mathbb{E}\left[w(Y(1,X_i)|Y(0,X_i))|X_i = X\right]\right)\,,
\end{aligned}$$

其中 $e(x) \overset{\text{定义}}{=} \mathbb{P}\left(T_i = 1|X_i = x\right)$。

2. 对于 $\pi(x) = 0$ 常数策略，$\psi_\pi = \psi_L$ 是失败比例，即主要干预相对于替代方案有害的可能性，其有效影响函数表示为；对于 $o = (X、\ T、\ Y)$:

$$\begin{aligned}
\phi_L(\mathcal{P})(o) = &-\psi_L(\mathcal{P}) + \mathbb{E}\left[w(Y(0,X_i)|Y(1,X_i))|X_i = X\right] \\
&+ \frac{T}{e(X)}\left(\mathbb{E}\left[w(Y(0,X_i)|Y)\right] - \mathbb{E}\left[w(Y(0,X_i)|Y(1,X_i))|X_i = X\right]\right) \\
&+ \frac{1-T}{1-e(X)}\left(\mathbb{E}\left[w(Y|Y(1,X_i))\right] - \mathbb{E}\left[w(Y(0,X_i)|Y(1,X_i))|X_i = X\right]\right)\,.
\end{aligned}$$

3. 对于一般策略 π， ψ_π 的有效影响函数为：

$$\phi_\pi(\mathcal{P})(o) = -\psi_\pi(\mathcal{P}) + \pi(x)(\phi_W(\mathcal{P})(o) + \psi_W(\mathcal{P})) + (1-\pi(x))(\phi_L(\mathcal{P})(o) + \psi_L(\mathcal{P}))$$

命题 *2* 的证明。我们有：

$$\begin{aligned}
\Psi_W(\mathcal{P}) &\stackrel{\text{def}}{=} \mathbb{E}\left[\mathbb{E}\left[w(Y_i(1,X_i)|Y_i)\right]|X_i, T_i = 0\right]] \\
&= \int \mathrm{d}\mathbb{P}_{X_i}(x)\mathrm{d}\mathbb{P}_{Y_i|X_i=x,T_i=0}(y)\mathrm{d}\mathbb{P}_{Y_i|X_i=x,T_i=1}(y')w(y'|y) \\
&= \int p(x)p(y|x,0)p(y'|x,1)w(y'|y)\mathrm{d}x\mathrm{d}y\mathrm{d}y' \\
&= \int p(x)\frac{p(x,0,y)}{p(x,0)}\frac{p(x,1,y')}{p(x,1)}w(y'|y)\mathrm{d}x\mathrm{d}y\mathrm{d}y' ,
\end{aligned}$$

其中， p 是相对于勒贝格测度的$\mathcal{P}$ 的密度（若$\mathrm{d}\mathcal{P}\ \mathrm{d}\lambda\ < \infty$），且该表达式可自然推广至一般情形。设 $h \in (0,1)$ （将取为无穷小）且$o = (X,T,Y)$。令$\mathcal{P}_h = (1-h)\mathcal{P}_h\delta_{X,T,Y}$为这两个分布的混合分布。我们有：

$$\Psi_W(\mathcal{P}_h) = \int p_h(x)\frac{p_h(x,0,y)}{p_h(x,0)}\frac{p_h(x,1,y')}{p_h(x,1)}w(y'|y)\mathrm{d}x\mathrm{d}y\mathrm{d}y' ,$$

其中， p_h 是$\mathcal{P}_h$的密度。注意，我们有 $p_h = (1-h)p + h\delta_{X,T,Y}$，因此：

$$\begin{aligned}
\Psi_W(\mathcal{P}_h) = (1-h)\Psi_W(\mathcal{P}) + o(h) &+ h\int p(y|X,0)p(y'|X,1)w(y'|y)\mathrm{d}y\mathrm{d}y' \\
&+ h\frac{1-T}{1-\pi(X)}\int p(y'|X,1)w(y'|Y)\mathrm{d}y' - h\frac{1-T}{1-\pi(X)}\int p(y|X,0)p(y'|X,1)w(y'|y)\mathrm{d}y\mathrm{d}y' \\
&+ h\frac{T}{\pi(X)}\int p(y|X,0)w(Y|y)\mathrm{d}y - h\frac{T}{\pi(X)}\int p(y|X,0)p(y'|X,1)w(y'|y)\mathrm{d}y\mathrm{d}y' .
\end{aligned}$$

由于 $\Psi_W(\mathcal{P}_h) = \Psi_W(\mathcal{P}) + h\Phi(\mathcal{P})(o) + o(h)$，我们可以如命题 1 中那样确定 Φ_W。对于Φ_L 和一般策略 Φ_π 的证明完全相同。 □

E 详细实验部分

E.1 合成模拟细节

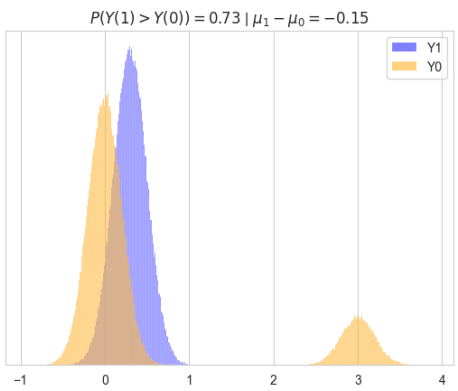
在合成实验中，我们探讨将 CPTE 方法应用于基于偏好的策略学习与评估所带来的益处。我们以连续型结局上的 PNS 为例，$\mathbb{E}\left[w(Y_i(1),\ Y_j(0))|X_i = X_j = x\right] = \mathbb{P}\left(Y_i(1) > Y_j(0)|X_i = X_j = x\right)$。

在某些情况下，基于 CATE 的策略足以优化 CPTE 目标。例如，由于伯努利分布由其期望值唯一确定，因此估计 CATE 足以学习二元 PNS 偏好。然而，复杂的潜在结果分布可能导致最优 CATE 策略与最优的基于偏好的策略（如 PNS）之间出现差异。下面我们给出一个真实案例。

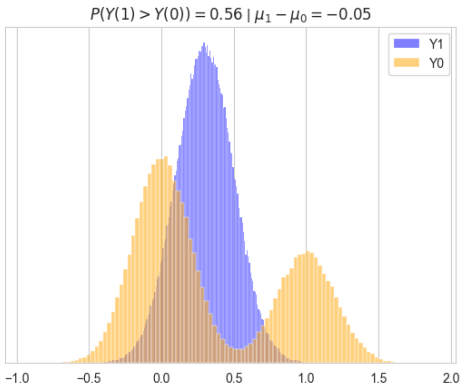
为生成实验数据，我们模拟了二元和连续特征的混合。潜在结果由线性效应修饰因子与复杂治疗效应组合生成。我们提出了两种情景：一种具有同质处理效应，另一种具有简单的异质处理效应。模拟满足假设 1 至 3。为检验 CPTE 方法与经典的基于 CATE 的策略，我们对复杂治疗效应进行模拟，使两种策略对每个个体的建议都不相同。当某个潜在结果分布具有重尾时，就可能出现这种情况。现实中的例子可能是某些高风险手术：它们能让特定

个体受益匪浅，却会对大多数患者造成伤害。在这种情况下，少数受益显著的患者可能会拉高平均值，使风险更高的手术看起来比更保守、让患者受益的可能性更高的方案更有益。CATE策略会建议进行风险手术；CPTE策略则不会。

我们的模拟通过从高斯分布中抽取一个潜在结果，并从双峰正态分布中抽取另一个潜在结果，来模拟这一情境。图 2a 和 2b 展示了两个潜在结果分布的示例，在这些示例中，基于 PNS 的策略与基于 CATE 的策略存在差异。还可以使用伽马分布和帕累托分布等重尾分布来模拟其他更微妙的示例。



(a) 差异较大



(b) 差异较小

图 2：潜在结果分布

在同质情境中，所有个体的潜在结果均从同一个潜在结果分布中抽取，并根据特征的线性组合进行调整。在异质性设置中，我们使用一个观测到的二元特征来决定将双峰分布分配给哪个潜在结果。更多细节请参见 E.1.1 节。

对于每次模拟，我们使用三种分布式学习范式——KNN、线性回归和随机森林——进行策略学习和评估。我们在一个留出集（N=10000）上评估这些方法；该留出集使用的随机种子未用于生成任何训练数据模拟。对于每个 N，我们将训练数据模拟 50 次。我们为每个$N \in [30,\ 100,\ 1000,\ 10000]$抽取一个训练集，并训练我们的分布估计方法和基线估计方法。利用它们对条件 PNS 的估计，我们评估三种策略规则。第一种是 OTR（5.2节），使用 $\delta(x) > 0$规则；CPTE 使用 $\delta(x) = q_W(x) - q_L(x)$，CATE 使用 $\delta(x) = \mathbb{E}\left[Y_i(1) - Y_i(0)|x\right]$。第二种策略规则是策略价值最大化（5.3节），它使用逻辑回归模型，在训练数据上学习$f:\ X \to \mathbf{I}(\delta(x) > 0)$。策略遵循预测结果。第三种策略规则同样使用逻辑回归模型来决定策略，但它基于 $\delta(x)$的 EIF 校正（公式 (13)）进行学习。

我们将采用分布估计的条件概率处理效应方法与采用条件平均处理效应的标准方法进行比较，后者作为基线。分布式 K 近邻（D-KNN）估计$\mathcal{P}_{Y_i|T_i,\ X_i}$，并与估计 $\mathbb{E}\left[Y_i|T_i,\ X_i\right]$的常规 K 近邻进行比较。同样，采用分位数回归的分布式线性回归（DLR）与常规岭回归进行比较；分布式随机森林（DRF）与常规随机森林进行比较。

E.1.1 合成数据生成过程

我们使用 8 个连续变量和两个二元变量生成特征 $X \in \mathbb{R}^{10}$。我们随机抽取线性系数，为潜在结果创建一个依赖于 X 的共享基线。在观察性情境中，我们选择三个特征（一个离散特征和两个连续特征）作为混杂因素

并使用线性变换创建用于治疗组选择的对数几率。我们确保存在重叠性，并将期望治疗率设为 0.5。

- 连续特征： $X_{cont} \sim \mathcal{N}(0, 1)$
- 离散特征： $X_{disc} \sim Ber(0.5)$
- 同质治疗分配： $T \sim Ber(0.5)$
- 异质治疗系数： $\beta_T \sim U(0.2, 0.6)$ 对于作为混杂因素的一个离散特征和两个连续特征，其他情况下为 0。
- 异质治疗分配： $T \sim Ber(X^T \beta_T)$
- 基线结果系数： $\beta_Y \sim U(0.1, 0.5)$ 适用于所有特征。
- 共享基线 Y： $Y_{baseline} = X^T \beta_Y$

我们为潜在结果空间定义了两种分布：双峰高斯混合和简单高斯分布。如图 2a 和图 2b 所示，双峰分布的期望值高于简单高斯分布。然而，简单高斯分布中的随机变量高于双峰分布中的随机变量的概率超过 0.5。我们令 $b \in [0, \ 1]$参数控制属于较低高斯分布的个体数量，以及 $1-b$ 属于较高高斯分布的个体数量。参考上文提到的风险手术示例， $1-b$ 表示从风险手术中获益显著的少数群体患者数量。

潜在结果参数:

- 共享方差： $\sigma = 0.2$
- 简单高斯参数： $\mu_s = 0.3$
- 简单高斯分布： $y_{simple} \sim \mathcal{N}(\mu_s, \ \sigma^2)$
- 高斯混合参数： $\mu_{m, \ low} = 0, \mu_{m, \ high} = 3, b = 0.85$
- 高斯混合： $y_{mixture} = b * \mathcal{N}(\mu_{m, \ low}, \sigma^2) + (1-b)\mathcal{N}(\mu_{m, \ high}, \sigma^2)$

我们通过将线性基线结果与这些分布相结合来生成潜在结果。在同质情境下,

$$Y(1) = y_{baseline} + y_{simple}$$

$$Y(0) = y_{baseline} + y_{mixture}$$

因此，对于所有个体，最优 PNS策略选择 T=1，而最优 CATE策略选择 T=0。

我们将其中一个二元特征记为 X_0。在异质性设置下,

$$Y(1) = y_{baseline} + \mathbf{I}(X_0 = 1) y_{simple} + \mathbf{I}(X_0 = 0) y_{mixture}$$

$$Y(0) = y_{baseline} + \mathbf{I}(X_0 = 0) y_{simple} + \mathbf{I}(X_0 = 1) y_{mixture}$$

因此，最优的 PNS 和 CATE 策略取决于二元特征 X_0。

E.1.2 实验详情

评估。 评估是在一个大型留出集（N=50,000）上进行的。 X 和y_{simple}， $y_{mixture}$使用所有训练集均未使用的随机种子生成。 β_y 的参数则使用所有数据集共享的随机种子生成。因此，评估数据和训练数据中的 $y_{baseline}$ 来自同一分布。

训练。 对于每个 $N \in \{30,\ 100,\ 1000,\ 10000\}$，训练数据均经过 50 次模拟。这使我们能够观察模型在规模呈指数增长的数据集上的表现。每次 X 和y_{simple}， $y_{mixture}$ 都是使用不同的随机种子生成的。

对于每种学习范式——KNN、线性回归和随机森林——我们都训练一个分布模型和一个基线模型。分布模型在假设 1 至 3 下估计潜在结果的分布，$\hat{} \ Y_i(1),\ \hat{}Y_i(0)|X_i = x$ 。分布式 K 近邻（DKNN）使用最近邻来近似潜在结果的分布。我们测试最近邻（k=1）以及按对数增长的 k 值。分布式线性回归器和森林回归器使用分位数近似逆累积分布函数，然后从近似的潜在结局分布中抽样 1000 次。我们将使用的样本数记为 S。基线学习器在假设 1 至 3 下估计潜在结果的期望值。这些估计使我们能够进行策略学习和评估。

策略通过三种方式学习：

- 代入最优治疗规则（5.2节）
- 代入式策略价值优化（5.3节）
- EIF策略价值优化（见第6节）

策略学习与策略评估对于每种学习范式，我们使用评估集上的预言机策略价值，将CPTE策略与CATE策略进行比较。这使我们能够了解哪种策略表现最佳。完整结果见图3，并将在下一节中分析。

策略评估通过使用每种学习范式在评估集上的估计值，并比较据此得到的最优策略价值来进行。CATE策略价值的计算方法是将估计的期望值代入偏好函数（定义1）。

Mo删除参数基线线性回归模型为岭回归，alpha=1。 0.
随机森林的参数取决于 N 的大小。这些参数是通过针对每个 N 进行超参数搜索得出的，搜索所用的随机种子未用于训练。这样，我们便能了解在不同训练集大小下哪些参数较为合理。

表 1：不同样本量下的随机森林回归器超参数设定 N。

条件为 N	最大深度	最大特征数	最小样本分割数	树的数量（$n_{\text{estimators}}$）
$N \geqslant 10000$	25	0.35	5	500
$100 < N < 10000$	15	0.50	7	400
$N \leqslant 100$	15	0.60	5	50

E.1.3 合成实验：详细结果

策略学习。如图3所示，分布式方法往往会收敛到最优策略价值。我们还注意到，以自助法95%置信区间表示的方差，对于分布式方法而言往往会逐渐减小。相比之下，基线学习器往往会收敛到错误的CATE策略，且方差往往更大。该模拟为线性模拟，因此，在所有设定和所有策略学习方法中，所有分布式线性回归（DLR）策略都能迅速收敛到最优策略。基线插件法和直接

策略价值优化收敛到最优CATE策略，这与最优CPTE策略恰恰相反。我们注意到，CATE的EIF PNS校正使得某些设定下的表现优于随机策略。分布式K近邻（DKNN）策略也会逐渐收敛到最优策略价值。尽管DKNN插件法不适用于线性模拟，但它能正确估计PNS，因此可以按照定理2所预期的对数速率学习到正确的策略。与DKNN不同，分布式随机森林（DRF）方法在异质性设置下难以克服模型设定错误的问题。虽然它不像基线随机森林那样收敛到CATE策略，但它难以学习正确的分布。因此，由于异质性均衡分布，插件估计始终接近随机策略。尽管存在这一不足，策略价值优化方法最终仍收敛到最优策略。

策略评估。 除了分布方法能够在复杂的潜在结果分布上学习策略之外，它们还能对诸如PNS之类的复杂偏好函数进行正确设定的策略评估。我们使用来自分布模型和基线模型的PNS估计，以PNS衡量最优策略。如图4所示，分布估计收敛于最优策略。正确设定的分布式线性回归迅速收敛到正确的策略价值。DKNN和DRF趋向于正确的策略价值。相比之下，使用期望潜在结果的基线估计则发散至0和1。我们注意到真实策略价值为0.72。

潜在结果之间的相关性。 我们还考察定理1中的假设被违反的情形，即必须观测到所有结构性治疗效应修饰因子。该假设不太可能成立；这些实验有助于我们理解偏离该假设的影响。我们通过引入无法由可用特征解释的潜在结果之间的相关性来检验这一点。延续手术示例，风险手术成功的患者可能具有某种遗传成分，该成分也会影响他们对更为保守处理的反应。

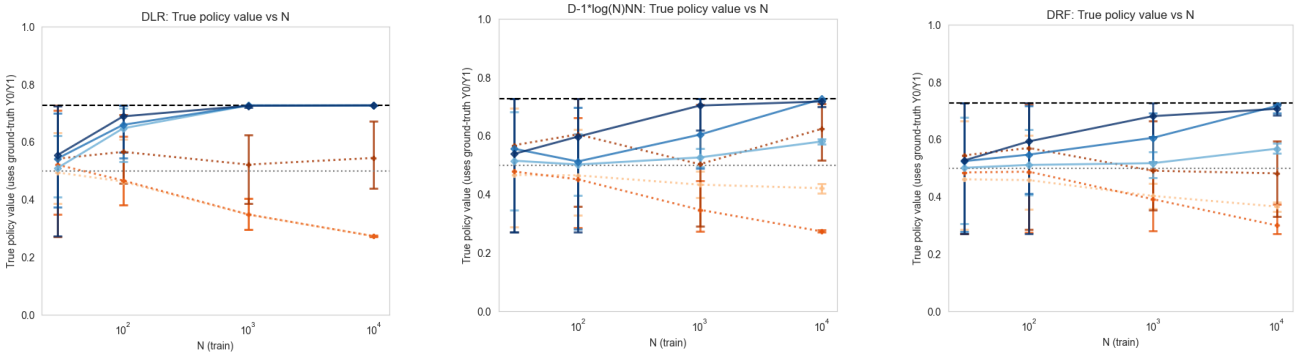
我们表明，在我们简单的合成环境中，策略学习对这一假设的违背似乎比策略评估更为稳健（图5）。对于策略学习，我们聚焦于在困难的异质性设置中表现成功的DKNN和分布式线性回归。我们在潜在结果之间加入了大量正相关和负相关，其经验皮尔逊相关系数分别为0.7和-0.7。

策略学习几乎不受影响。图6的可视化展示了相关性仅影响跨越决策边界（红线）的一小部分总体（蓝色同心圆）。然而，策略价值估计受到显著影响。我们在图7中展示了使用正确设定的线性模型进行的策略评估。在存在相关性时，策略价值收敛到错误的策略价值。它们收敛到无相关性时的策略价值。

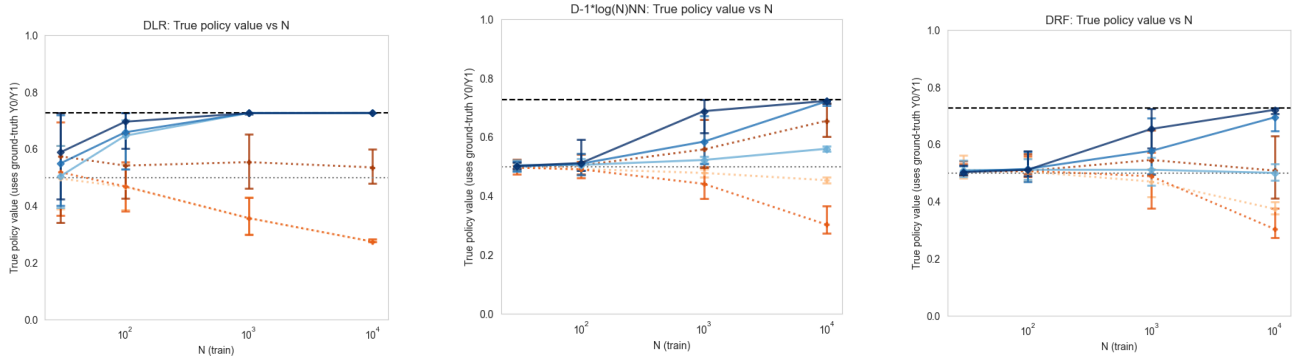
E.2 半合成模拟细节

验证因果推断方法很困难，因为验证所需的反事实真实值从未被观测到。这一现象被称为因果推断的根本挑战[19]。自然的解决方案是在理想情形下验证方法，例如随机对照试验。随机对照试验能为条件平均处理效应估计提供无偏估计，因为它们在构造上是无混杂的。虽然条件个体处理效应估计同样受益于随机对照试验固有的可识别性，但它们在统计上仍然存在问题。均值可以在较小的子层上得到合理估计。然而，估计CPTE所需的条件分布需要更多数据。这一点尤其成立，因为CPTE特别适用于具有复杂条件潜在结果分布的任务，正如我们在上述模拟中对PNS所展示的那样。

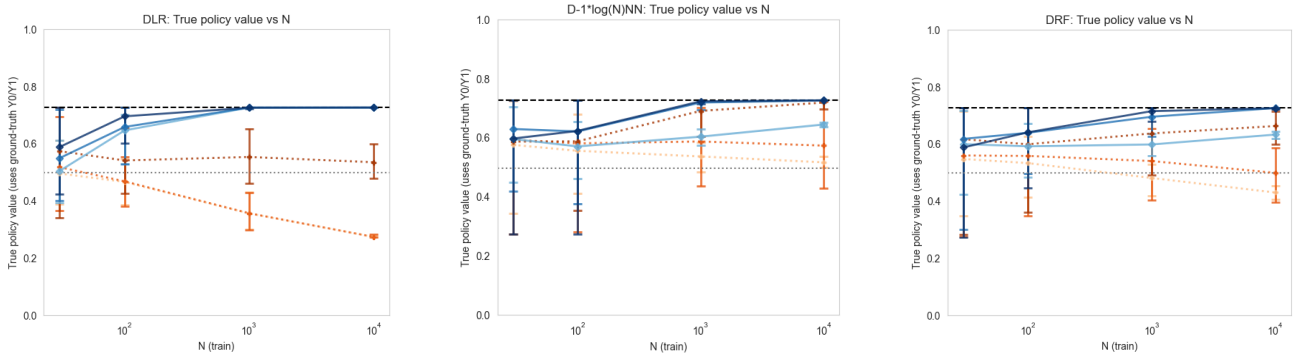
因此，为了进一步验证我们的方法，我们转向半合成实验。我们以田纳西STAR研究 [59] 为起点。田纳西学生/教师学业成绩比（STAR）研究是一项随机试验，旨在衡量班级规模对学生考试的影响



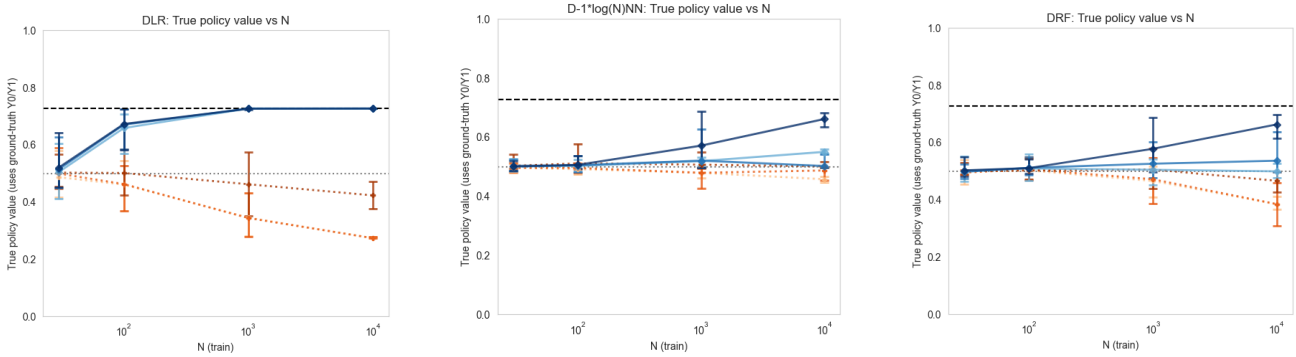
(a) DLR (b) DKNN (c) DRF



(d) DLR (e) DKNN (f) DRF

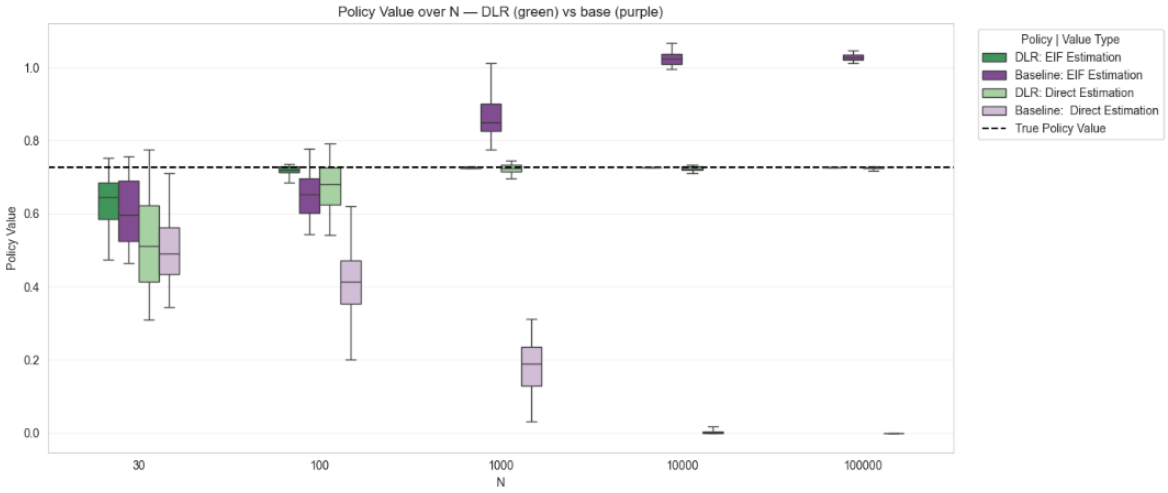


(g) DLR (h) DKNN (i) DRF

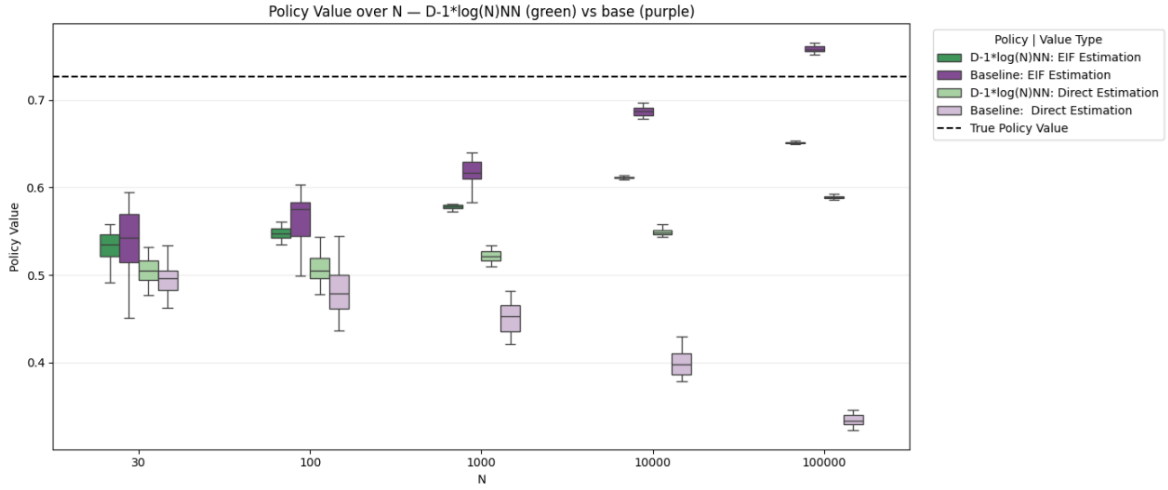


(j) DLR (k) DKNN (l) DRF

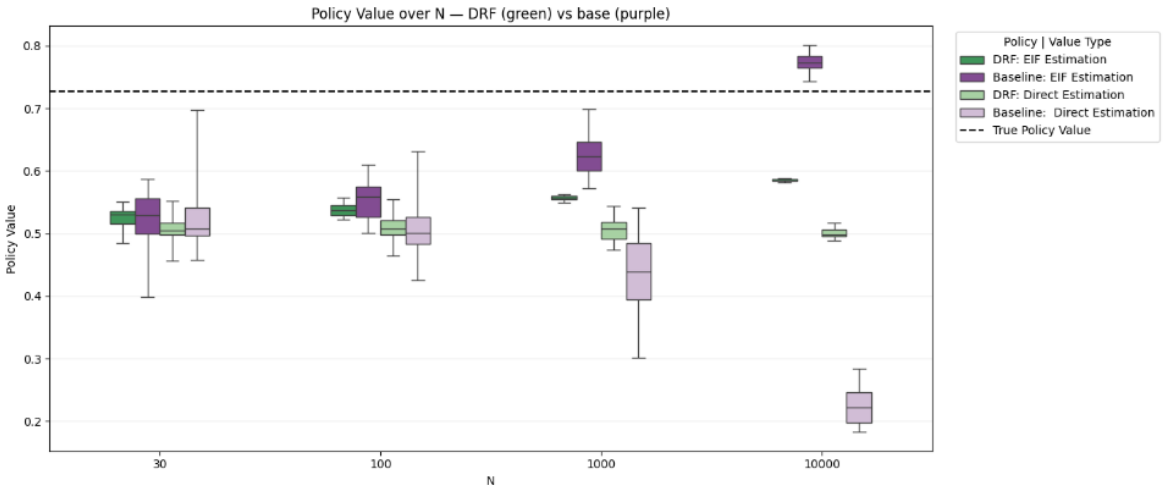
图3：无相关性时基于PNS策略价值的策略学习。(a–c) 同质随机对照试验设定；(d–f) 异质性随机对照试验环境；(g-i) 同质观察性设定；(j-l) 异质性观察性情境。**图例** 与图1a相同。



(a) DLR

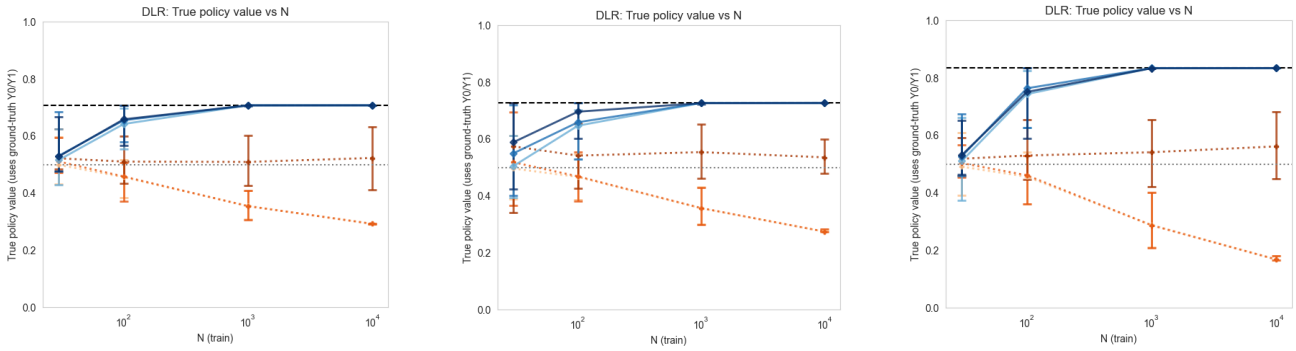


(b) DKNN

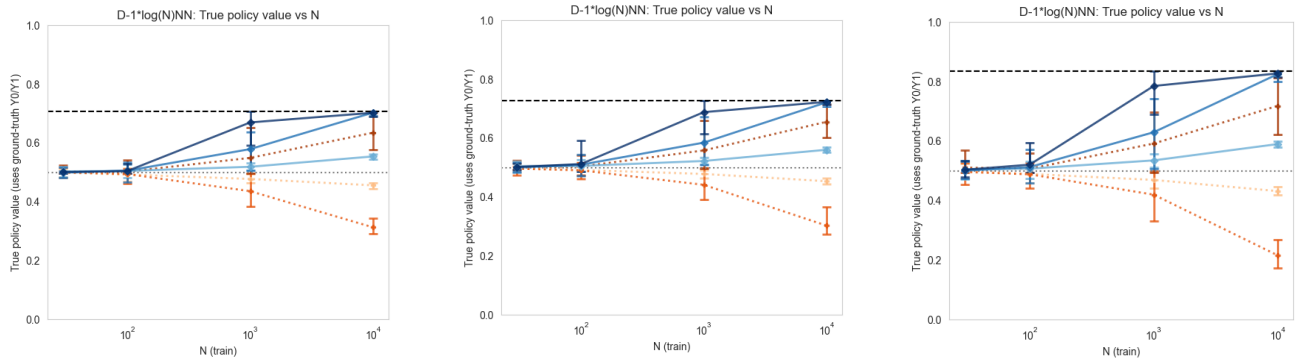


(c) DRF

图4：合成异质性随机对照试验环境中最优策略的策略评估



(a) DLR：-0.7　(b) DLR：0　(c) DLR：0.7



(d) DKNN：-0.7　(e) DKNN：0　(f) DKNN：0.7

图5：合成的异质性随机对照试验环境中，具有相关潜在结果的策略学习。（a–c）线性回归；（d–f）DKNN。**图例** 同图1a。

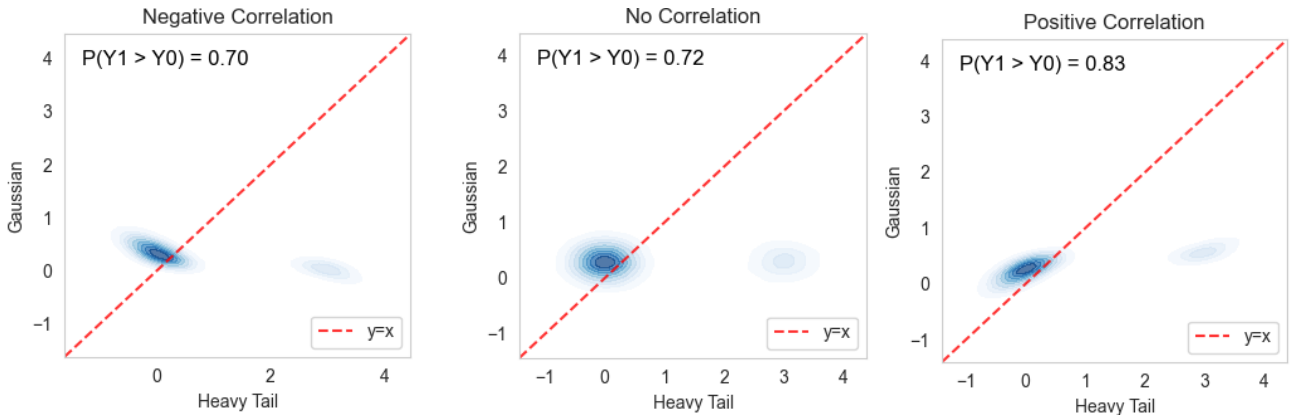


图6：可视化具有相关潜在结果的 PNS、$P(Y1 > Y0)$、$Y1$ 和 $Y0$。在我们的实验中，$Y1$ 服从高斯分布（y 轴），而 $Y0$ 具有重尾（x 轴）。

分数，即数学、阅读和听力成绩。这项研究于1985年启动，追踪了K-3年级的学生，并在每个学年结束时记录他们的成绩。学生被随机分配到三种班级类型之一：小班（13-17名学生）、普通班（22-25名学生）和配有全职助教的普通班（22-25名学生）。每所学校至少设有一个每种类型的班级。[2]。我们将方法应用于“获胜”函数，即一种偏好函数（定义1）；该函数通常被汇总为各种“获胜”统计量 [12]。“获胜”函数刻画了对层级结局的偏好。

我们研究一个主要二元结局和一个次要连续型结局。这种情形在医学文献中很常见：主要结局通常是生存（以二元结局或事件发生时间的形式衡量），次要结局则是血压等连续型结局。我们的主要结局是实验失访。我们考察了研究中纳入的幼儿园学生，并将失访定义为在一年级时未能在研究中找到的学生。我们的次要结局是学生的数学成绩。我们将“获胜”定义为在一年级时仍在研究中且数学成绩较高的学生。

剔除没有数学分数的学生后，小班有 1,762 名学生，常规班有 2,032 名，配有全职助教的常规班有 2,077 名。另有 454 名学生因缺少数学分数而被剔除。由于“获胜”函数只定义在两种处理之间，我们将常规班的两个处理组进行合并。尽管这可能违反稳定单元处理值假设，但研究结果之一是，全职助教的影响并不显著 [2, 29]。

如前所述，在随机对照试验中通过经验方法计算条件“获胜”函数在统计上存在问题，因为许多较小的子层会导致“获胜”函数估计不佳，尽管这些估计在因果意义上是无偏的。因此，我们生成合成结局。我们力求复现原始结局分布。为此，我们在结局数据上训练多层感知机，然后使用 MC-dropout 从该模型中抽样 [16]，从而得到条件分布的贝叶斯近似。

随后，我们将这些条件分布估计值作为最优策略及其价值的真实值。我们采用与合成实验类似的方式评估方法，训练分布模型和基线模型。然后，我们将这些模型的预测作为“获胜”函数的插件估计量，并使用与合成实验相同的策略学习方法预测策略。

在合成实验中，我们展示了对于复杂的潜在结局分布，基于 PNS 的策略相较于基于均值的策略所具有的优势。即使没有这种复杂性，“获胜”函数的优势也十分明显。由于“获胜”函数由层级结局构成，成功的、基于“获胜”的策略能够区分不同个体：对于某些个体，次要结局带来的收益超过主要结果中的风险。估计结果期望值的基线学习器可能只会优先考虑主要结果。在二元情形下，只有在估计的期望值相等时，才会考虑次要结局。

我们使用原始特征和模拟结果，展示了半合成 STAR 数据上的结果。与合成实验相比，这种更贴近现实的模拟中，偏好接近决策边界，因此策略学习更加困难。

从图 8 可以看出，随着训练数据规模增大，分布方法最能学到最优策略。在图中，我们展示了预言机策略价值（黑色）和随机策略（灰色）的结果。我们还加入了最优主要结果策略作为基准（棕色），也就是说，如果策略仅依据主要结果，而非“获胜”偏好函数来制定，其策略价值会是多少。我们发现，所有基线插件策略（黄色）甚至都无法达到最优主要结果策略的水平。分布策略（蓝色）和基线的直接策略价值优化策略则超过了最优主要结果策略。这些结果表明，分布估计量在优化 CPTE 函数方面具有强大能力。相比之下，插件基线策略仅略微超过 0.5 的“偏好”价值。这表明，CATE 估计值在“获胜”统计量这类 CPTE 结果上的表现不佳。

我们还在半合成设置中对相关潜在结果进行了实验。我们的实验表明，我们的模型和基线模型对这一假设违背并不敏感——

定理1的条件。参见图 9。

E.2.1 数据收集与处理

创建队列我们使用 STAR 数据集作为半合成实验的起点，该数据集见于 AER R 包 [26]。我们选取幼儿园学生，并通过检查哪些学生出现在一年级队列中来构建二元结果。数学结果数据缺失（454 名学生）。缺失的数学结果数据被剔除。我们将学生分为两个干预组：小班和常规班。后一个处理组包括两组：没有助教的常规班，以及配有全职助教的常规班。尽管这可能违背稳定单元处理值假设（SUTVA），但研究结果表明，全职助教的作用并不显著 [2, 29]。最终形成两组：小班教室组（1,762 名学生）和普通教室组（4,109 名学生）。

特征 用于生成模拟结果和进行训练的相关特征包括：学生性别、学生族裔、教师性别、教师族裔、教学年限、教师最高学位、教师所处的职业阶段（例如见习教师、试用期、一级等）、学校所在地（乡村／中心城区／城市／郊区），以及学校是否提供免费／减价午餐。最后一个变量是低收入家庭背景的指示变量 [2]。除教学年限外，所有变量均为分类变量。分类变量采用独热编码，缺失值在所有类别中均赋值为零。教学年限中的缺失值则赋为中位数。

表 2：缺失变量

变量	变量类型	缺失值	预处理
Math	结果	454	已删除
学生族裔	特征	1	所有类别均赋值为零
免费/减价午餐	特征	17	所有类别均赋值为零
学位	特征	21	所有类别均赋值为零
职业阶梯	特征	541	所有类别均赋值为零
教师族裔	特征	60	所有类别均赋值为零
教学年限	特征	21	中位数

结果模拟 针对每个干预组和结果分别训练了一个多层感知机。我们汇总了各模型的超参数和评估指标。所有模型在训练时均采用了 10% 的随机失活率

表 3：深度模型的汇总

模型	隐藏维度	训练轮次	批量大小	学习率	Brier分数	均方根误差（标准差）
小型 – 随机失活	[64, 64]	35	128	5×10^{-4}	0.17	–
常规 – 随机失活	[64, 64]	34	128	5×10^{-4}	0.19	–
小型 – 数学分数	[64, 64, 64]	100	32	1×10^{-4}	–	48.5 (49.5)
常规 – 数学分数	[64, 64, 64]	100	32	1×10^{-4}	–	45.1（46.7）

模型训练完成后，我们使用 MC-dropout [16] 进行抽样，丢弃率为 5%。对于每个个体，我们针对每个干预组抽取 1000 个潜在结果。这些模拟构成了一个困难且并非简单的策略学习问题。大多数个体模拟得到的“胜出”偏好函数值介于 0.4 和 0.6 之间。我们比较主要结果和次要结局对应的最优策略：

表4：最优策略比较

策略	给予较小者治疗	与胜出策略的一致性	胜出策略价值
胜出（随机失活、数学）策略	77%	100%	0.59
辍学策略	67%	70%	0.54
数学成绩策略	75%	92%	0.58

如表4所示，优先考虑次要结局的策略接近最优策略。这是由于二元结局上的治疗效应并不显著。然而，这给在“胜出”设定下学习最优策略带来了困难。“获胜”的层级排序迫使策略首先考虑主要结果。

训练与评估 训练与评估过程与合成模拟相似。主要区别在于加入了一个二元主要结果。二元结果完全由其伯努利参数（即期望值）确定。对于主要结果，我们无需使用分布方法。因此，我们在训练数据上进行模型选择，并对所有分布方法和基线方法使用相同的模型。我们发现，使用 11 个邻居的K近邻分类器可获得最佳 Brier分数（0.17）。我们的样本量也受幼儿园队列规模的限制。我们将数据划分为一个包含 1,500 名学生（25%）的留出集，其余 4,403 名学生作为训练数据。我们随机抽取 $N \in [30,\ 100,\ 1000,\ 4403]$ 条训练数据，并用这些数据训练估计模型及其对应的策略学习模型。对于每个模型和 N，我们重复这一过程五十次。

E.2.2 相关性

我们还在半合成研究中探讨了潜在结果之间的相关性所产生的影响。为此，我们采用 Iman-Conover 方法 [22]，该方法常用于在模拟中诱导变量之间的相关性 [20, 51, 14]。此方法无分布假设，且能保留边际分布。它不会改变数值，而是对数值重新排序，从而诱导出对目标相关性的近似。我们使用此方法诱导每种结果的潜在结果之间的相关性。因此，在三个实验中，边际潜在结果分布完全相同。对于每个实验，我们从每个个体的潜在结果中抽取一个样本，并根据该个体所属的治疗组确定观测结果。

我们评估了潜在结果之间存在正相关或负相关时模型的表现，结果见表5。对于正相关，每个结果的目标相关性设为 0.5。二元潜在结果和连续型潜在结局所诱导的相关性分别为 0.28 和 0.47。对于负相关，每个结果的目标相关性设为 -0.5。二元潜在结果和连续型潜在结局所诱导的相关性分别为 -0.22 和 -0.41。

表5：比较最优策略

相关性（诱导）	策略	给予较小者治疗	与获胜策略的一致性	获胜策
-0.5 (-0.22/-0.41)	胜出（随机失活、数学）策略	73%	100%	
	辍学策略	67%	62%	
	数学成绩策略	75%	98%	
0	胜出（随机失活、数学）策略	77%	100%	
	辍学策略	67%	70%	
	数学成绩策略	75%	92%	
0.5 (0.28/0.47)	胜出（随机失活、数学）策略	74%	100%	
	辍学策略	67%	63%	
	数学成绩策略	75%	99%	

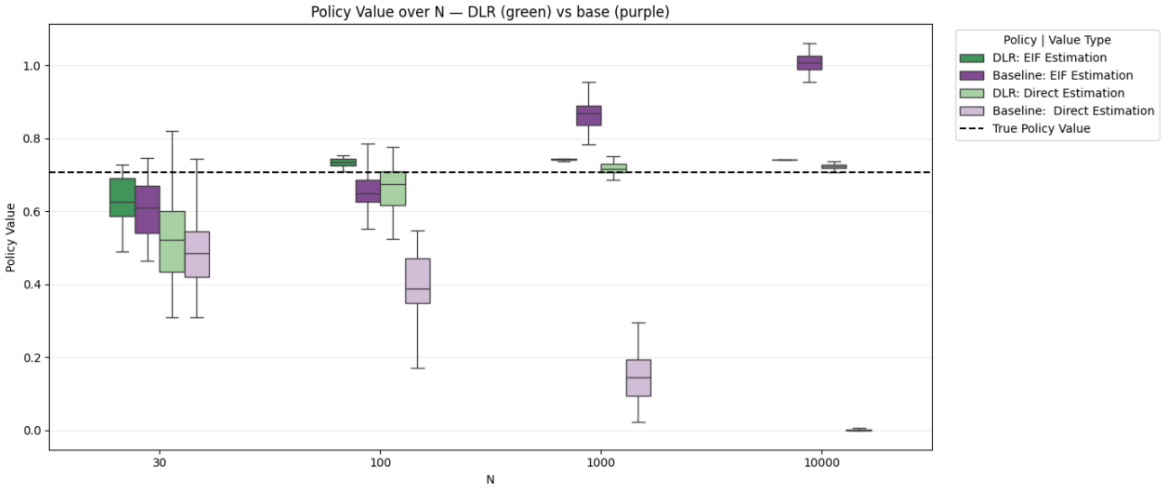
与合成实验类似，相关性会使最优策略略有变化（8%），而正相关会提高最优策略价值。我们在图 9 中展示，我们的模型仅受到轻微影响。由于正相关会使个体的“获胜”函数更加极端，我们发现，随着模型更好地学会正确的策略，其策略价值会显著提升。

E.2.3 KNN 中邻居数量的影响

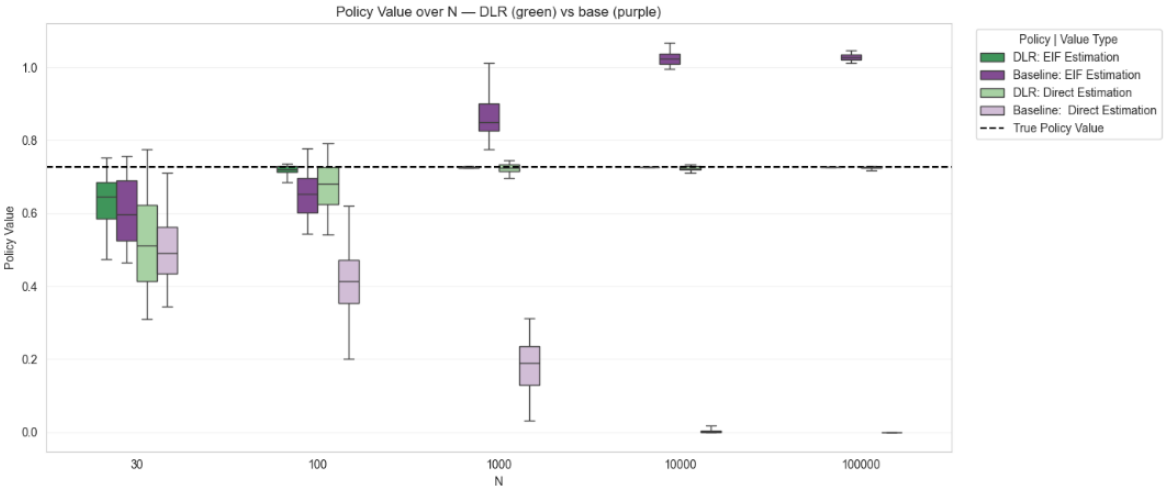
图 10 展示了在策略学习中增加更多邻居的影响。正如我们的理论所预测的，插件式 OTR（5.2节）需要 $k \to \infty$ 才能收敛。事实上，对于 $k = 1$ ，插件法的表现与基线插件法一样差（几乎与随机猜测一样差），而增大 k会带来更高的策略价值，且在更高的 k下收敛更快。然而，颇为有趣的是，策略价值优化方法（A.3节和5.3节）并非如此，因为价值是在整体人群上取平均的。事实上，增加更多邻居似乎对收敛影响不大。

E.2.4 策略一致性 - 优化“获胜”对比主要结果对比次要结局

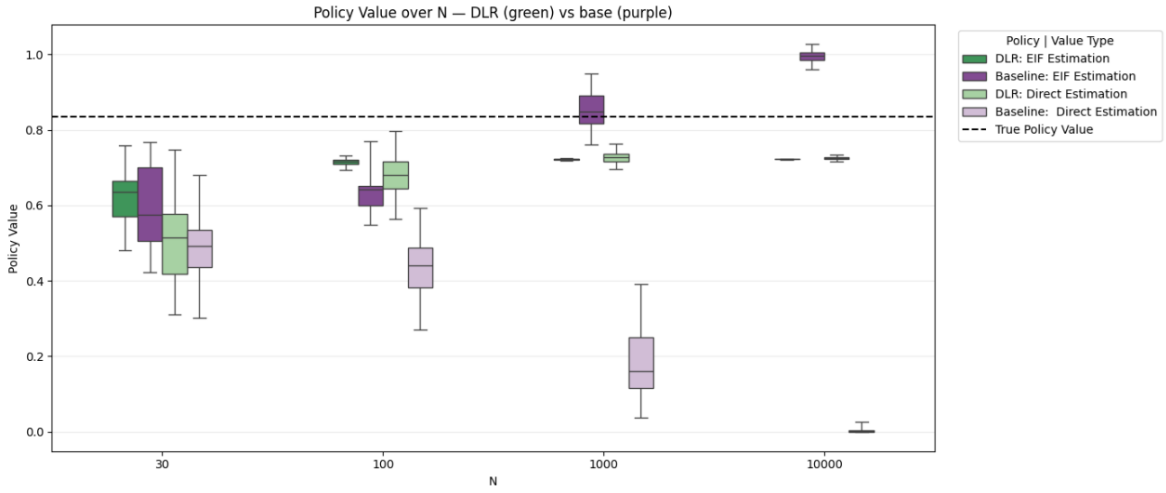
最后，我们比较策略一致性，即所学策略与给定预言机策略之间的一致性百分比；该预言机策略为以下之一：*(i)* 使用层级结局、借助预言机信息学习得到的最优策略，*(ii)* 仅优化主要结果、借助预言机信息学习得到的最优策略，以及*(iii)* 仅优化次要结局、借助预言机信息学习得到的最优策略。结果报告于图 11。我们展示了 DLR 方法的结果。这些结果对 KNN 和 DRF 也同样成立。



(a) DLR - 负相关

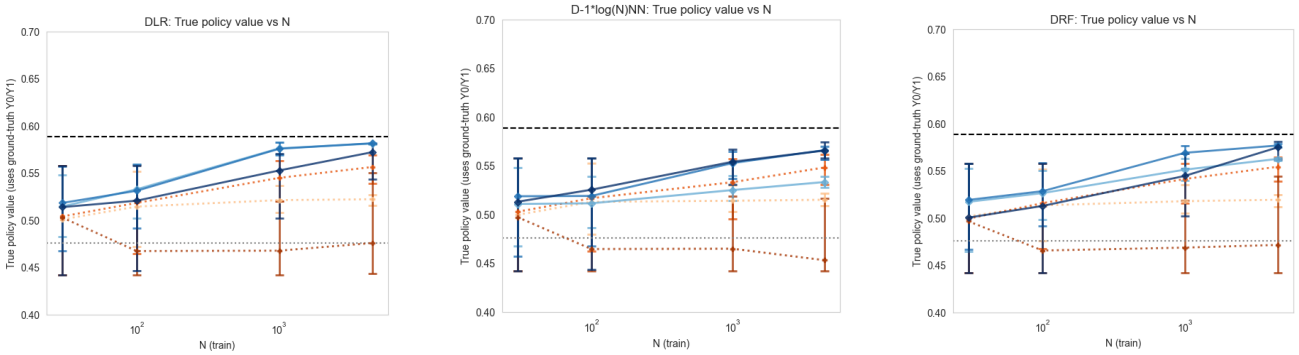


(b) DLR



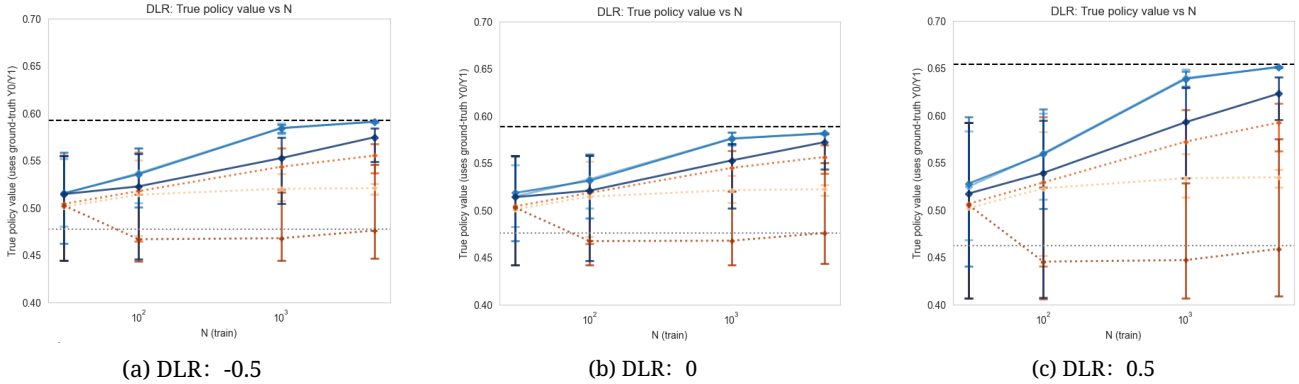
(c) DLR - 正相关

图7：半合成设置中相关性下最优策略的策略评估

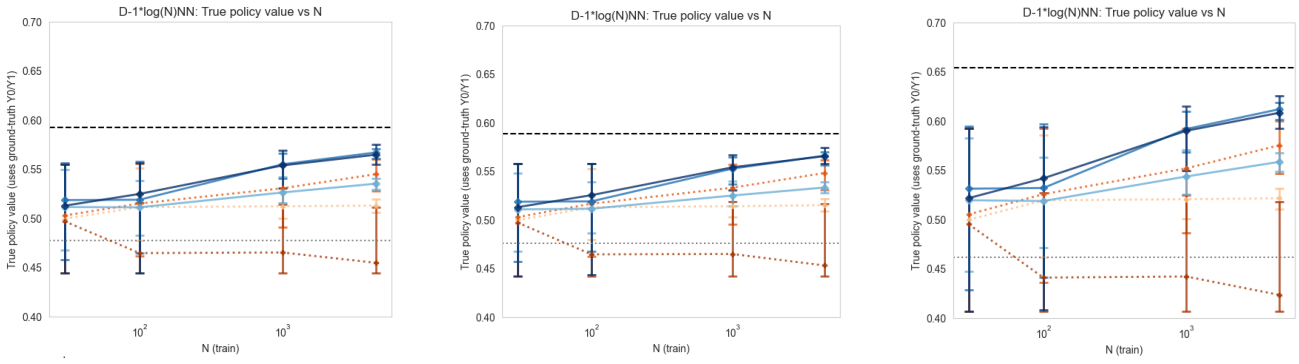


(a) DLR (b) DKNN (c) DRF

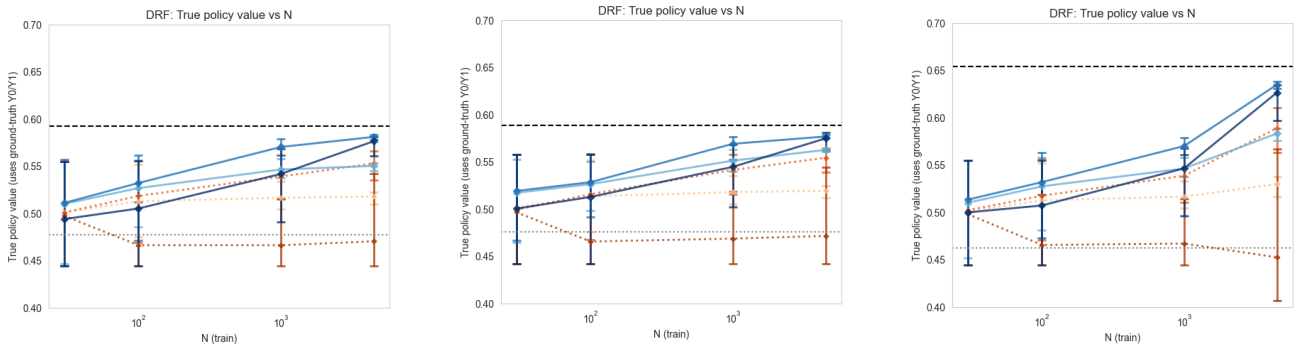
图8：策略学习 中使用半合成设置中的获胜策略价值。**图例**同图1a .



(a) DLR：-0.5 (b) DLR：0 (c) DLR：0.5

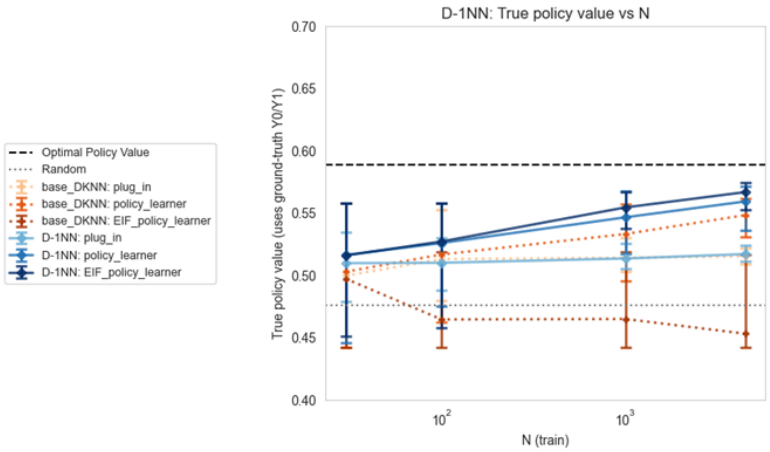


(d) DKNN：-0.5 (e) DKNN：0 (f) DKNN：0.5

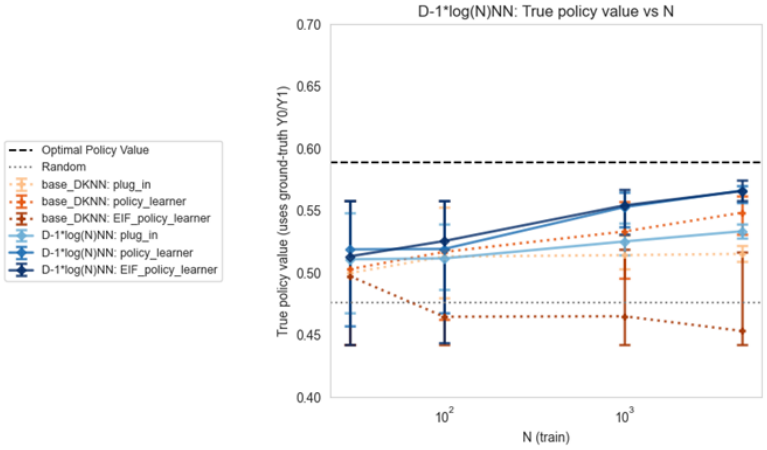


(g) DRF：-0.5 (h) DRF：0 (i) DRF：0.5

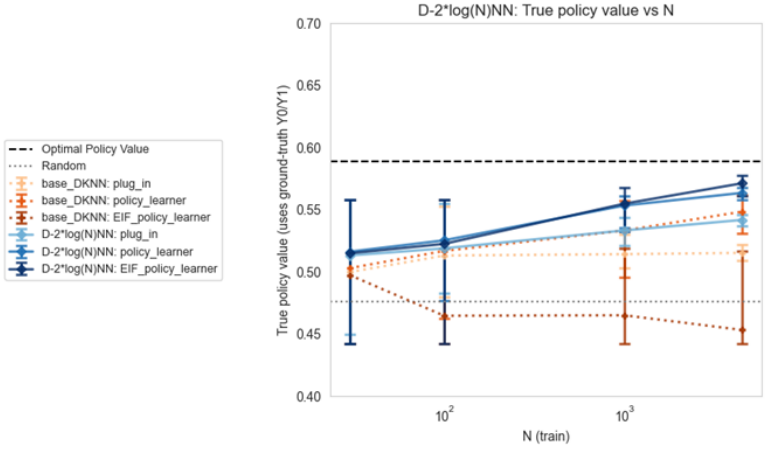
图9：半合成设置中潜在结果之间存在相关性时的策略学习。(a–c) 线性回归；(d–f) DKNN；(g–i) DRF。**图例**同图1a。



(a) $k = 1$ 邻居

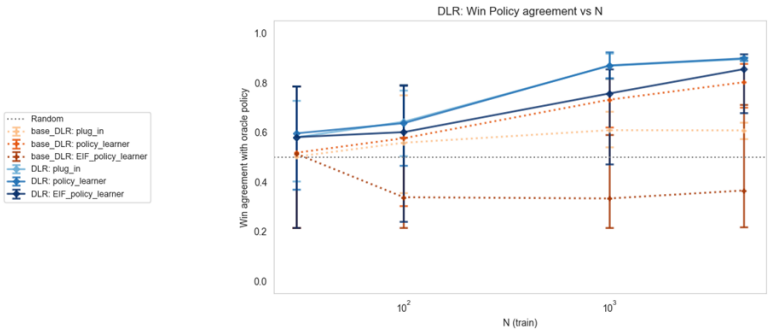


(b) k =log(n) 邻居

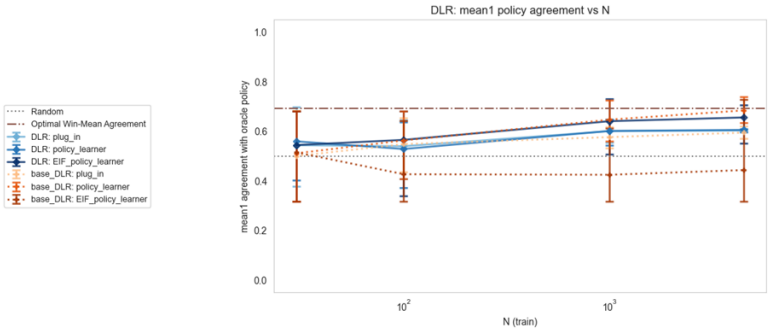


(c) $k = 2\log(n)$ 邻居

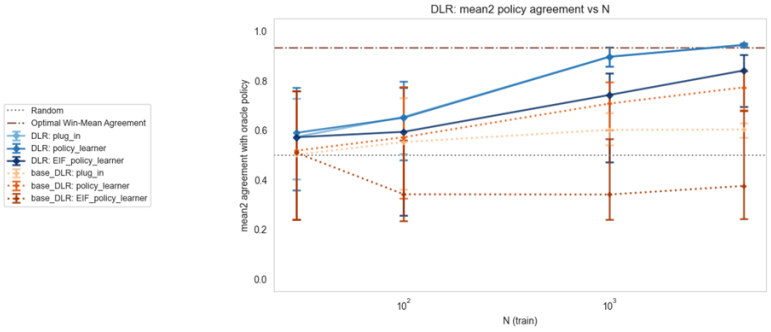
图 10： k–NN（ $k = 1$、 $k = \log(n)$ 和 $k = 2\log(n)$ ），在无相关性的半合成实验中



(a) 与按获胜情况进行优化的预言机策略的一致性



(b) 与按主要结果进行优化的预言机策略的一致性



(c) 与按次要结果进行优化的预言机策略的一致性

图 11：学习策略与不同预言机策略的一致性