因果剃刀：因果发现中简洁性假设的比较研究*

Wai-yin (
Wayne)Lam哲学系 卡内基梅
隆大学 匹兹堡，宾夕法尼亚州，
美国

arXiv:2302.10331v3 [cs.LG] 7 Aug 2023

摘要

进行因果发现时，必须对真实因果机制如何对应于底层联合概率分布作出假设。在本文中，这些假设被称为因果剃刀。我们回顾了文献中出现的众多因果剃刀，并对它们进行了全面的逻辑比较。特别是，我们仔细审视了一个不受欢迎的因果剃刀，即多项因果模型中的参数最小性，以及它与其他经过充分研究的因果剃刀之间的逻辑关系。我们的逻辑结果使得为基于评分的因果搜索算法选择合理评分准则时面临两难。

1 引言

科学推断的一个基本目标是发现因果结构。如何基于非实验数据实现这一目标，是一个引人注目的问题，尤其是在数据丰富的时代。近几十年来，因果发现研究蓬勃发展，这主要得益于数据的爆炸式增长，以及许多能够降低计算复杂性的机器学习技术的兴起。传统统计工具旨在恢复非因果关联，而因果发现则致力于从观察数据中推断给定系统内变量之间因果关系的过程。

因果发现程序也称为因果搜索算法，通常将观察数据中的统计信息与描绘因果关系的图形表示联系起来。一种广泛使用的图形工具是因果有向无环图（DAG），其中每条有向边都表示从一个变量流向另一个变量的直接因果影响。通过假设真实的因果机制可以用因果 DAG 表示，因果搜索算法主要旨在从真实情况生成的观察数据中恢复因果信息。

给定一组测量变量 V 上具有联合概率分布$\mathcal{P}$的观测数据集，V 上的每个有向无环图 $\mathcal{G}$ 都是一种假设，其目标是通过图结构特征来解释 $\mathcal{P}$ ，尤其是通过 Pearl [1988]著名提出的 d-分离来解释。对 $\mathcal{P}$的解释主要指 $\mathcal{P}$所蕴含的条件独立关系（CI）集合，这是研究人员挖掘潜在因果信息的重点。因此，因果发现中的一项直观任务是缩小假设空间，以便能够从真实的因果机制中恢复大量因果信息。然而，假设空间的规模会随 V 的大小呈指数级增长。因此，一个关键的研究问题是如何排除较差的假设。

结构发现若不做任何假设，几乎无法实现。因果发现也不例外。只有假设真实的

*本文是作者学位论文《因果剃刀与因果搜索算法》的一部分 [Lam， 2023]。

因果机制（记作 $\mathcal{G}^*$）满足某些期望的特征，才能淘汰较差的假设；不具备这些特征的有向无环图会被视为较差。受 Reichenbach 的共同原因原则启发， [1956], 一个广为人知的假设是因果马尔可夫条件（CMC）：它要求由 $\mathcal{G}^*$ 通过 d-分离所蕴含的条件独立关系，是联合概率分布 $\mathcal{P}$中成立的条件独立关系的子集。循此思路，违反 CMC 的有向无环图（即非马尔可夫有向无环图）属于不可接受的假设，应予以拒绝。

尽管 CMC 得到了广泛认可，但仅凭 CMC，我们仍有大量假设需要考虑。例如，每个完全有向无环图（即所有顶点两两之间都由有向边相连）都显然满足 CMC，因为它通过 d-分离不蕴含任何条件独立关系。因此，要获得关于 $\mathcal{G}^*$ 的更多实质性信息，就需要更强的假设。文献中广泛讨论的一种候选假设是因果忠实性条件（CFC），简称忠实性。忠实性采用 CMC 的逆命题，假设 $\mathcal{P}$ 中成立的条件独立关系集合是 $\mathcal{G}^*$ 在图结构上所蕴含的条件独立关系集合的子集。尽管忠实性具有很强的逻辑强度，能够大幅缩小假设空间，但即使满足 CFC， $\mathcal{G}^*$ 也未必可识别。方向不同的有向无环图在图结构上可能蕴含相同的条件独立关系集合。按照标准记法，这些有向无环图属于同一个马尔可夫等价类（MEC）。一项重要的理论结果是，当真实因果模型（$\mathcal{G}^*$， $\mathcal{P}$）满足 CFC 时， $\mathcal{G}^*$ 的 MEC（即真实MEC）是可识别的。

值得肯定的是，正如 Spirtes 等人 [2000], 所示，对于广泛研究的许多概率分布，CFC 在总体层面上的违反集合的勒贝格测度为零。尽管如此，在实践中，非忠实性并不会以极小概率消失。有时，科学模型的参数值是在长期动态平衡过程中演变而来的，因此容易导致非忠实性[（Andersen， 2013]）。对于有限样本，CFC 的违反可能是近似的（而非精确的）：统计检验会将依赖关系误判为独立关系（即近似独立关系） [（Zhang 和 Spirtes， 2008]）。因此，为了避免这种情况，人们或许会倾向于采用强因果忠实性条件，该条件假设联合分布中几乎不存在任何未由 CMC 蕴含的 CI[（Kalisch 和 B¨uhlmann， 2007]）。然而， [Uhler 等人 2013] 令人信服地表明，近似不忠实性经常出现。这一观察促使研究人员反思：尽管 CFC 在总体层面具有良好的数学性质，在样本层面假设 CFC 是否合理。

下面的例子说明了这一现象。假设我们希望分析以下五个变量之间的因果关系：服用避孕药（C）、糖尿病（D）、吸烟（K）、妊娠（P）和中风（T）。设图1中的 $\mathcal{G}^*$ 描绘了真实的数据生成因果机制，我们从该机制中收集有限样本。我们观察到，T直接受到其他四个变量中每一个变量的影响。例如，C对T存在正向因果影响，这是由于服用避孕药后发生中风的概率略有增加。另一方面，C还通过P对T产生间接但负向的因果影响。其含义是，服用避孕药会降低妊娠的概率，而妊娠会提高中风的概率。假设C对T的直接（正向）和间接（负向）因果影响大小相对接近并趋于抵消，也就是说，尽管C与T在概率上具有依赖关系，但C对T的概率只造成了可忽略的变化。由于抽样误差，我们从有限样本中错误地得出了C与T边际独立的结论。C与T之间的近似独立性并未在图中表示为 $\mathcal{G}^*$ （因为二者之间存在邻接关系）。因此， $\mathcal{G}^*$ 不忠实于抽样分布，并且表现出近似违反CFC的情况。

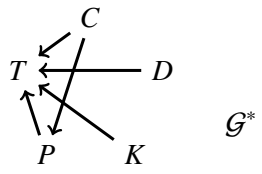


图1：一个解释服用避孕药（C）、糖尿病（D）、吸烟（K）、妊娠（P）与中风（T）之间因果关系的DAG假设。

历史上，PC算法 [Spirtes 和 Glymour， 1991]）以及GES算法 [Chickering， 2002]已被证明能够在CFC下识别真实MEC。然而，在非忠实性条件下，它们可能无法完成这一任务。例如，由于C与T近似独立，PC算法会错误地判定二者不相邻。因此，近似不忠实性问题推动了这样一个研究方向：研究严格弱于CFC的假设，同时仍能保证真实MEC可识别。这些假设包括Ramsey 等人 [2006], 提出的忠实性的受限形式，以及Raskutti 和 Uhler [2018]提出的最稀疏马尔可夫表示（SMR）假设。文献中还研究了许多类似的理论假设。例如，Zhang [2013] 讨论了两个极小性条件，即Pearl极小性和SGS-极小性，认为真实因果模型应满足这两个条件。加上CFC，他将这三个条件称为马尔可夫因果模型的奥卡姆剃刀。显然，这一术语源自奥卡姆的威廉提出的简约律，该定律优先考虑简洁性：在相互竞争的假设中，应优先选择更简单的那个。鉴于近几十年来，与因果发现相关的众多简洁性假设已得到研究，我们借用类似的术语，将所有这些假设统称为因果剃刀。

目前已有十多种因果剃刀被提出，但对它们进行全面的逻辑比较仍付之阙如。一个可能的例外是 [Forster 等人， 2020] 他们将少数几种因果剃刀与简约性进行了比较：简约性是最稀疏马尔可夫表示（SMR）的一种较弱变体，要求排除并非最稀疏的马尔可夫DAG。例如，考虑图2中的DAG假设 $\mathcal{G}'$ ，据称它与图1中的 $\mathcal{G}^*$ 一样也具有马尔可夫性。按照简约性原则， $\mathcal{G}'$ 应当被排除，因为 $\mathcal{G}^*$ 包含的边严格少于 $\mathcal{G}'$。

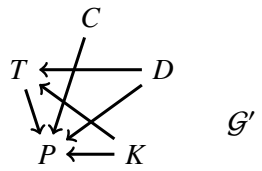


图2：满足CMC的另一种DAG假设。

表面上看，简约性在简洁性方面具有直观吸引力。与计算DAG中的边数类似，心理学实验研究（例如，Bonawitz 和 Lombrozo [2011]）发现了积极证据，表明人们会根据因果解释中所援引的原因数量，来判断因果假设的简洁性。此外，某些因果搜索算法（例如，Raskutti 和 Uhler [2018], Solus 等人 [2021]）声称通过偏好更稀疏的马尔可夫DAG来识别真实MEC。

采用以图结构稀疏性定义的因果剃刀，其理论动机之一源自广泛研究的线性高斯因果模型。这类模型建立了有向无环图中的有向边与线性系数之间优雅的双射，而线性系数是定义因果模型所用的参数。因此，简约性的直观吸引力可以说植根于一个更基本的概念：因果模型中的参数数量。

统计学和计算机科学领域普遍采用参数计数。这有助于衡量统计/学习模型的统计复杂度；一般而言，参数较少的假设不易出现过拟合问题 [MacKay, 1992]。然而，因果发现文献中尚未有人认真尝试分析参数最小性，更不用说探讨它与简约性等其他因果剃刀之间的关系。

上述论点促使我们对因果剃刀展开更具百科全书式的分析。通过以 CMC 和 CFC 为界，包容性地列出一系列因果剃刀，我们将从逻辑强度的角度对其进行审视。另一方面，我们表明，对因果剃刀进行全面的逻辑分析虽然很容易被忽略，却能得出重要启示，补充日益受到重视的因果搜索算法发展。一个例子是 [Lam 等人近期的研究， 2022] 该研究

表明 CFC 与一种特定形式的 Pearl极小性在逻辑上等价。这一逻辑发现可用于揭示一种因果搜索算法的局限性，而其开发者并未意识到这一点。例如，Solus 等人 [2021]曾认为，他们的 GSP算法在非忠实性情形下仍然正确。然而，当 CFC 不成立时，这种逻辑等价性表明，他们的算法可能会陷入一个 Pearl极小DAG；在他们的框架中，该图被视为次优。这一例子验证了对因果剃刀进行审慎比较研究的实际意义。

另一方面，我们聚焦于多项式因果模型中简约性与参数最小性之间的比较。我们将展示一个例子，以证明这两种因果剃刀之间的逻辑独立性，即更稀疏的DAG（例如图1中的 $\mathcal{G}^*$ ）可能比更稠密的DAG（例如图2中的 $\mathcal{G}'$ ）包含更多参数。尽管当我们掌握关于变量的背景知识时， $\mathcal{G}'$ 看似不太可信，但在没有背景知识或情境信息的情况下，尚不清楚 $\mathcal{G}^*$ 和 $\mathcal{G}'$ 哪一个是更可取的假设。此外，这种逻辑独立性与因果搜索算法的实际实现密切交织，尤其是GES和GSP等基于评分的算法。面对近似不忠实性，我们应采用边数作为评分准则，以偏好更稀疏的马尔可夫DAG，还是采用一种偏好参数更少模型的度量（例如Schwarz的贝叶斯评分准则 [1978]）？ 这一困境促使算法用户反思：在进行稳健的因果发现时，他们究竟默认采用了哪一种因果剃刀。

在深入探讨更多细节之前，有必要明确说明关于因果剃刀的几点补充意见。首先，尽管“剃刀”一词体现了奥卡姆剃刀的精神，但本文将研究的并非每一种因果剃刀都明确体现了简洁性。例如，方向忠实性 [Ramsey 等人， 2006] 要求有向无环图中的顶点三元组不与某种特定的条件独立关系相关联。很难看出，这一假设如何能够与理论上的节俭性联系起来。因此，本文中的因果剃刀主要是指压缩假设空间这一行为。

其次，这种压缩行为应当从功能角度加以理解。根据 Woodward 所倡导的因果的功能论 [2021], , 因果推理关注的是行动者出于特定目标和目的，想要利用因果概念做什么。因此，需要结合一个定义明确的期望目标来研究因果剃刀的压缩行为，以便告知我们压缩达到何种程度才是必要且充分的。按照通常的做法，本文将重点放在识别真实MEC上。尽管如此，文献中仍存在不同的形式化期望目标。LiNGAM算法 [Shimizu等人， 2006] 并不寻求识别真实MEC，而保守PC算法 [Ramsey 等人， 2006] 旨在识别一个比真实MEC更宽泛的有向无环图类。按照 Lin和 Zhang 所采用的学习理论方法 [， 2020],在假定某种特定因果剃刀的情况下，应当探究所能达到的最高期望目标是什么。对期望目标的研究也可以从实践角度加以理解。假定较强因果剃刀的算法，其运行速度可能远快于假定较弱因果剃刀的竞争算法。当缺乏计算资源时，研究人员可以通过将后者作为期望目标，合理地选择以牺牲准确性换取效率。本文并不坚持对期望目标进行客观研究，而是倾向于提供一份可能性菜单，供研究人员自由选择。

最后，参数假设也可用于缩小可能假设的范围，因此也应相应地视为因果剃刀。然而，参数假设通常与 CFC 和简约性等非参数定义的结构因果剃刀无关。尽管如此，参数最小性是一个显著的例外，因为要有效枚举参数，就必须采用参数假设。本文仅在讨论参数最小性时探讨参数假设。感兴趣的读者不妨进一步研究参数假设与结构因果剃刀之间的相互作用。

本文结构如下。第2节首先介绍一些基本术语。第3节回顾文献中研究过的十一种结构因果剃刀，随后在第4节对其进行全面的逻辑分析。第5节讨论参数最小性及其更强变体，以及它们在多项式情境下与结构因果剃刀之间的逻辑关系。第6节开始探讨一种算法层面的

困境，这是我们的逻辑分析所揭示的。第7节将给出总结性评述。证明和示例的构造细节将留在附录中。

2 基本术语

本文中，我们始终用斜体字母表示变量（例如，X_1,Y），用粗体字母表示变量集合（例如，X）。下文将回顾有向无环图和联合概率分布相关的一些标准符号和定义。

在测量变量集合 $\mathbf{V} = \{X_1, \ldots$ 上的有向图 $\mathcal{G}$ 由 m 个顶点 $\mathbf{v} = \{1, \ldots, m\}$ 组成，其中每个顶点 $i \in \mathbf{v}$ 与变量 $X_i \in \mathbf{V}$ 相关联，且 $\mathcal{G}$ 中的每条边都是有向的，形式为 $j \to k$。我们关注简单图，其中没有顶点指向自身的边，并且每对顶点之间至多有一条边。对于任意顶点 j、$k \in \mathbf{v}$，如果 j 和 k 之间存在一条有向边（即，$j \to k$ 或 $k \to j$），则称它们在 $\mathcal{G}$ 中相邻。令 $E(\mathcal{G})$ 表示 $\mathcal{G}$ 中的边集合。

给定一个有向图 $\mathcal{G}$，路径 p 是由顶点组成的序列 $\langle p_1, \ldots, p_k \rangle$ 对于某个 $k \geq 2$，其中 p_i 和 p_{i+1} 在 $\mathcal{G}$ 中相邻，对每个 $1 \leq i < k$ 均成立。若对每个 $1 \leq i < k$ 都有 $(p_i \to p_{i+1}) \in E(\mathcal{G})$ 成立，则称路径 p 为 $\mathcal{G}$ 中的一条有向路径（从 p_1 到 p_k）。有向无环图（DAG）是一种有向图，其中没有顶点存在一条指向自身的有向路径。

给定变量集 V，记 DAG(v) 为定义在 V 上的所有可能 DAG 的集合。下面介绍一些标准的图表示法。给定一个 DAG $\mathcal{G} \in$ DAG(v) 和一个顶点 $i \in \mathbf{v}$:

- $pa(i, \mathcal{G}) := \{j \in \mathbf{v} : (j \to i) \in E(\mathcal{G})\}$：i 在 $\mathcal{G}$ 中的父节点集合；
- $ch(i, \mathcal{G}) := \{j \in \mathbf{v} : (i \to j) \in E(\mathcal{G})\}$：i 在 $\mathcal{G}$ 中的子节点集合；
- $ne(i, \mathcal{G}) := pa(i, \mathcal{G}) \cup ch(i, \mathcal{G})$：$\mathcal{G}$中 i 的邻居节点集合；
- $an(i, \mathcal{G}) :=$ $pa(i, \mathcal{G})$ 的传递闭包与 {i} 的并集：$\mathcal{G}$中 i 的祖先节点集合；
- $de(i, \mathcal{G}) :=$ $ch(i, \mathcal{G})$ 的传递闭包与 {i} 的并集：$\mathcal{G}$中 i 的后代节点集合；
- $nd(i, \mathcal{G}) := \mathbf{v} \setminus de(i, \mathcal{G})$：$\mathcal{G}$中 i 的非后代节点集合。

$\mathcal{G}$的骨架记作 $sk(\mathcal{G})$，是 $\mathcal{G}$中邻接关系（无序顶点对）的集合。图示而言，当 $\mathcal{G}$是有向无环图时，$sk(\mathcal{G})$ 表示 $\mathcal{G}$ 去除 $\mathcal{G}$中所有箭头后得到的无向图。

对于 $\in \mathbf{v}$ 中任意三个两两不同的顶点 i、j、k，如果 (i, j)、(j, k) $\in$属于 $sk(\mathcal{G})$，但 (i, k) $\notin$不属于 $sk(\mathcal{G})$，则称 (i, j, k) 是无屏蔽的；如果它们在 $\mathcal{G}$中两两相邻，则称 (i, j, k) 是有屏蔽的（或构成一个三角形）。给定一条路径 $p = \langle p_1, \ldots,$ 在有向无环图 $\mathcal{G}$中，若 $1 < i < k$，则$(p_{i-1} \to p_i)$、$(p_{i+1} \to p_i) \in$属于 $E(\mathcal{G})$ 时，p_i是路径 p 上的碰撞点，否则为非碰撞点。

对于任意 $i, j \in \mathbf{v}$ 和任意 $k \subseteq \mathbf{v} \setminus \{i, j\}$，若在 $\mathcal{G}$ 中存在一条连接 i 和 j 的路径 p，且 p 上没有非碰撞点属于 k，并且 p 上的每个碰撞点 l 或 l 的某个后代属于 k，则在 $\mathcal{G}$ 中，给定 k 时 i 和 j 是 d-连接的。若给定 k 时 i 和 j 不是 d-连接的，则在 $\mathcal{G}$ 中，给定 k 时 i 和 j 是 d-分离的。对于任意互不相交的顶点子集 i、j、$k \subseteq \mathbf{v}$，若对于每个 $i \in$ i 和每个 $j \in$ j，在 $\mathcal{G}$ 中 k 都将 i 和 j d-分离，则在 $\mathcal{G}$ 中，给定 k 时 i 和 j 是 d-分离的，记作 $i \perp_{\mathcal{G}} j \mid k$。

另一方面，考虑定义在 V 上的联合概率分布 $\mathcal{P}$。对于 V 的任意子集 X、$Y \subseteq$，给定 Y 时 X 的条件概率分布记为 $\mathcal{P}(X \mid Y)$。对于 V 中任意两两不相交的变量子集 X、Y、$Z \subseteq$，如果 $\mathcal{P}(X \mid Y, Z) = \mathcal{P}(X \mid Z)$，则称在给定 Z 时 X 与 Y 条件独立，并记作 $X \perp\!\!\!\perp_{\mathcal{P}} Y \mid Z$，表示 $\mathcal{P}$ 中 X 与 Y 在给定 Z 时的条件独立性（CI）。当 $X = \{X\}$ 和 $Y = \{Y\}$ 为单元素集合时，我们记作 $X \perp\!\!\!\perp_{\mathcal{P}} Y \mid Z$（而不记作 $\{X\} \perp\!\!\!\perp_{\mathcal{P}} \{Y\} \mid Z$）。当 $Z = \varnothing$ 时，我们记作 $X \perp\!\!\!\perp_{\mathcal{P}} Y$（而不记作 $X \perp\!\!\!\perp_{\mathcal{P}} Y \mid Z$）。[1]

[1] $X \perp\!\!\!\perp_{\mathcal{P}} Y$ 在文献中通常称为边际独立性。不过，为方便起见，所有边际独立关系都标记为条件独立关系。

遵循 [Studen ́y 的做法， 2005], 我们使用独立性模型作为工具，将联合分布中成立的条件独立关系与有向无环图通过 d-分离所蕴含的条件独立关系并列比较。首先，我们将所有可能的条件独立关系定义为：

$$\mathtt{I}(\mathbf{V}) := \{\langle \mathbf{X}, \mathbf{Y} \,|\, \mathbf{Z}\rangle : \mathbf{X}, \mathbf{Y}, \mathbf{Z} \text{ are pairwise disjoint subsets of } \mathbf{V}\}.$$

联合分布 $\mathcal{P}$ （定义在 V 上）的独立性模型为

$$\mathtt{I}(\mathcal{P}) = \{\langle \mathbf{X}, \mathbf{Y} \,|\, \mathbf{Z}\rangle \in \mathtt{I}(\mathbf{V}) : \mathbf{X} \perp\!\!\!\perp_{\mathcal{P}} \mathbf{Y} \,|\, \mathbf{Z}\}$$

有向无环图 $\mathcal{G}$ （定义在 V 上）的独立性模型为

$$\mathtt{I}(\mathcal{G}) = \{\langle \mathbf{X_j}, \mathbf{X_k} \,|\, \mathbf{X_l}\rangle \in \mathtt{I}(\mathbf{V}) : \mathbf{j} \perp_{\mathcal{G}} \mathbf{k} \,|\, \mathbf{l}\}$$

其中，对每个 $\mathbf{i} \subseteq \mathtt{v}$， $\mathtt{X_i} = \{\mathtt{X}_j\}_{j\in\mathbf{i}}$。与上文介绍的简写类似，当 X {和 {Y}是单元素集合时，我们将 〈X, Y |Z〉记作.....，并将 〈X, Y〉 记作.....，其中 Z = ∅。

如果 $\mathrm{I}(\mathcal{G}_1) = \mathrm{I}(\mathcal{G}_2)$，则两个有向无环图 $\mathcal{G}_1$ 和 $\mathcal{G}_2$ 马尔可夫等价。有向无环图 $\mathcal{G}$的马尔可夫等价类记作 MEC($\mathcal{G}$)，定义为所有与 $\mathcal{G}$马尔可夫等价的有向无环图所构成的集合。两个有向无环图可以不同，但马尔可夫等价。一个简单的例子是 2 个顶点的情况：此时 $\mathrm{E}(\mathcal{G}_1) = \{1 \to 2\}$， $\mathrm{E}(\mathcal{G}_2) = \{2 \to 1\}$， 而 $\mathrm{I}(\mathcal{G}_1) = \mathrm{I}(\mathcal{G}_2) = \emptyset$。正如 Verma 和 Pearl [1988], 所证明的，两个有向无环图马尔可夫等价，当且仅当它们具有相同的邻接关系和非屏蔽碰撞点。每个马尔可夫等价类都可以用唯一的完全部分有向无环图（CPDAG）以图形方式表示；该图是一种由有向边和无向边组成的混合图。 [Meek, 1995] 证明了一组定向规则足以从有向无环图的骨架和非屏蔽碰撞点中得到唯一的 CPDAG。

因果模型是一个二元组（$\mathcal{G}$,$\mathcal{P}$），其中 $\mathcal{G}$ 是一个有向无环图， $\mathcal{P}$ 是定义在同一组变量 V 上的联合概率分布。我们将 $\mathcal{G}^*$ 记为真实数据生成有向无环图，因此（$\mathcal{G}^*$,$\mathcal{P}$）是真实因果模型，我们假定它始终存在。

接下来，我们用 param($\mathcal{G}$,$\mathcal{P}$) 定义因果模型（$\mathcal{G}$,$\mathcal{P}$）的参数集合。为了确定因果模型的参数集合，必须对 $\mathcal{P}$ 作出参数假设。因此，对于非参数模型，若我们不假定联合分布 $\mathcal{P}$ 属于任何特定的参数分布族，那么 param($\mathcal{G}$,$\mathcal{P}$) 就是一个未定义的概念。下文中，若底层联合分布 $\mathcal{P}$ 可从上下文中明确得知，我们将 param($\mathcal{G}$) 作为简写。本文将重点讨论两类因果模型：结构方程模型（SEMs）和多项因果模型。

定义在 $\mathrm{V} = \{\mathrm{X}_1, ..., \mathrm{X}_m\}$ 上的结构方程模型（$\mathcal{G}$, ϱ, $\mathcal{F}$），其中 $\varrho = \{\varrho_1, ..., \varrho_m\}$ 是一组噪声项，$\mathcal{F} = \{\mathrm{f}_1, ..., \mathrm{f}_m\}$ 是一组结构方程，且对每个 $1 \le \mathrm{i} \le \mathrm{m}$ 均满足以下条件。

$$X_i = f_i(\mathrm{pa}(i, \mathcal{G}), \epsilon_i) \tag{1}$$

令 $\mathcal{P}_{\boldsymbol{\epsilon}}$ 为噪声项的联合概率分布。给定 $\mathcal{F}$中的结构方程，只要能够用 ϱ_i 唯一解出 X_i，就会在 V 上诱导出一个联合分布 $\mathcal{P}$ 。文献中常研究的一类结构方程模型是线性高斯因果模型，其中 ϱ 服从多元高斯分布，且每个 $\mathrm{f}_i \in \mathcal{F}$ 都是线性方程。具体而言，每个变量 $\mathrm{X}_i \in \mathrm{V}$ 都由以下方程支配：

$$X_i = \left(\sum_{X_j \in \mathbf{X}_{\mathrm{pa}(i,\mathcal{G})}} \theta_{ji} X_j \right) + \epsilon_i \tag{2}$$

其中 0， $\theta_{ji} \in \boldsymbol{\theta}_{\mathcal{G}}$ 仅当 $\mathrm{j} \in \mathrm{pa(i, \mathcal{G})}$ 时成立，且每个 $\varrho_i \in \varrho$ 都服从高斯分布。一种广泛采用的假设是，假定 ϱ 中的噪声项相互独立。[2] 在这一假设下，我们称

[2]通过要求 ϱ 联合独立，可以用 X_i 在其父节点 $\mathrm{X}_{\mathrm{pa}(i,\mathcal{G})}$ 给定时的条件均值和偏方差，估计每个 $\varrho_i \in \varrho$ 的均值和方差。

$\text{param}(\mathcal{G}) = \boldsymbol{\theta}_{\mathcal{G}}$（不涉及 ϱ）。鉴于 $\boldsymbol{\theta}_{\mathcal{G}}$ 与 $\text{E}(\mathcal{G})$ 之间存在双射，我们得到一个优美的等式：$|\text{param}(\mathcal{G})| = |\text{E}(\mathcal{G})|$，对每个线性因果高斯模型 $(\mathcal{G}, \mathcal{P})$ 都成立。

另一方面，在一个多项因果模型（$\mathcal{G}, \mathcal{P}$）中，$\mathcal{P}$ 是一个联合多项分布，其中每个 $\text{X}_i \in \text{V}$ 都是一个离散的 k 值变量，其取值范围由 $\text{range}(\text{X}_i) = \{0, 1, ..., \text{k} - 1\}$其中 $\text{k} \geq 2$。令 r(i) 表示 X_i 的取值范围的基数（即，$\text{r(i)} := |\text{range}(\text{X}_i)|$）。因此，我们称 V 的取值范围为 $\langle \text{r}(1), \text{r}(2), ..., \text{r}(|\text{v}|)|\rangle$。我们进一步将任意变量子集 $\text{X}_\mathbf{i} \subseteq \text{V}$ 的 range($\text{X}_\mathbf{i}$) 定义为 $\{\text{range}(\text{X}_j)\}_{j \in \mathbf{i}}$ 的笛卡尔积。类似地，我们将 $\text{r}(\cdot)$ 扩展到任意变量子集 $\text{X}_\mathbf{i} \subseteq \text{V}$，如下所示：

$$\text{r}(\mathbf{i}) := \begin{cases} \prod_{i \in \mathbf{i}} \text{r}(i) & \text{if } \mathbf{i} \neq \emptyset \\ 1 & \text{otherwise.} \end{cases}$$

对于每个 $\text{i} \in \text{v}$，令 $\boldsymbol{\theta}_{i,\mathcal{G}}$ 表示在 $\mathcal{G}$ 中计算 $\mathcal{P}(\text{X}_i | \text{X}^{\text{pa}(i,\mathcal{G})})$ 所需的 i 的参数集合。$\boldsymbol{\theta}_{i,\mathcal{G}}$ 中的每个参数都对应于给定特定取值配置 $\text{s} \in \text{range}(\text{X}^{\text{pa}(i,\mathcal{G})})$ 时，X_i 取某一特定值的概率。由于对于每个 $\text{s} \in \text{range}(\text{X}^{\text{pa}(i,\mathcal{G})})$，$\text{P}_{j \in range(X_i)} \mathcal{P}(\text{X}_i = \text{j} \mid \text{X}^{\text{pa}(i,\mathcal{G})} = \text{s}) = 1$，计算 $\mathcal{P}(\text{X}_i | \text{X}^{\text{pa}(i,\mathcal{G})} = \text{s})$ 的自由度（即逻辑独立值的最大数量）为 $\text{r(i)} - 1$.[3] 因此，我们有

$$|\boldsymbol{\theta}_{i,\mathcal{G}}| := (\text{r}(i) - 1) \times \text{r}(\text{pa}(i, \mathcal{G})). \tag{3}$$

因此，多项因果模型 $(\mathcal{G}, \mathcal{P})$ 的参数集合为 $\text{param}(\mathcal{G}) = \cup_{i \in \mathbf{v}} \boldsymbol{\theta}_{i,\ \mathcal{G}} = \boldsymbol{\theta}_{\mathcal{G}}$满足

$$|\text{param}(\mathcal{G})| = \sum_{i \in \mathbf{v}} |\boldsymbol{\theta}_{i,\mathcal{G}}| = \sum_{i \in \mathbf{v}} (\text{r}(i) - 1) \times \text{r}(\text{pa}(i, \mathcal{G})). \tag{4}$$

再来看图1和图2中的 $\mathcal{G}^*$ 和 $\mathcal{G}'$。假设所有变量均为二元，读者可以验证 $|\text{param}(\mathcal{G}^*)| = 21$ 和 $|\text{param}(\mathcal{G}')| = 23$。与线性高斯情形不同，对于每个多项因果模型 $(\mathcal{G}, \mathcal{P})$，都有 $|\text{param}(\mathcal{G})| > |\text{E}(\mathcal{G})|$ 成立。

基本符号介绍到此为止。下一节中，我们将利用这些符号阐述本文的核心，即因果剃刀。

3 结构因果剃刀

如第1节所述，因果发现的一般目标是从联合概率分布 $\mathcal{P}$中还原真实有向无环图 $\mathcal{G}^*$ 的因果信息。这也可以表述为逐步缩小假设空间 DAG(v)，直到只剩下所需的有向无环图类。要实现这种缩小，必然需要对真实因果模型 $(\mathcal{G}^*, \mathcal{P})$ 作出假设，从而排除不符合这些假设的有向无环图。本文将这些假设统称为因果剃刀。

粗略来说，因果剃刀可分为三类，尽管这些类别之间可能有所重叠。首先，可以对真实因果模型的联合概率分布 $\mathcal{P}$ 施加参数假设，从而将与 $\mathcal{P}$ 不相容的某些有向无环图排除在假设空间之外。例如，假设$(\mathcal{G}^*, \mathcal{P})$是一个线性高斯因果模型，则 $\mathcal{P}$ 中成立的条件独立关系只能由 $\mathcal{G}^*$ 中的 d-分离关系，或由 $\mathcal{G}^*$中路径抵消所导致的线性系数偶然抵消来解释。[4] 例如，将图1中的例子表示为线性高斯模型时，服用避孕药（C）与中风（T）之间的不忠实独立性源于

[3]由于 $\text{P}_{j \in \text{属于范围}(X_i)} \mathcal{P}(\text{X}_i = \text{j} \mid \text{X}^{\text{pa}(i,\mathcal{G})} = \text{s}) = 1$，$\mathcal{P}(\text{X}_i = \text{k} \mid \text{X}^{\text{pa}(i,\mathcal{G})} = \text{s})$ 在已知每个 $\text{j} \in$ 属于范围$(\text{X}_i) \setminus \{\text{k}\}$ 时，便可由 $\mathcal{P}(\text{X}_i = \text{j} \mid \text{X}^{\text{pa}(i,\mathcal{G})} = \text{s})$ 确定。[4]关于路径抵消的深入理论讨论，[请参见 Weinberger 2017]。

从 C 到 T 的两条有向路径之间的抵消效应。另一类因果剃刀具有算法性质。因果剃刀所规定的有向无环图集合由特定因果搜索算法的输出决定。例如，Solus 等人 [提出的最稀疏边排列（ESP）假设 2021]，以及 Lam 等人 [提出的 GRaSP 剃刀， 2022] 都属于这一类。

尽管如此，文献中更受关注的因果剃刀属于另一类。一般而言，其定义并不诉诸特定的参数特征或算法特征。它们关注的是明确定义的图结构特征（例如边），或由图结构特征所蕴含的特征（例如由 d-分离所蕴含的条件独立关系）。我们将这类因果剃刀称为结构因果剃刀。诚然，这种启发式分类并不精确。但这种有意保留的模糊性可以使我们接下来的比较分析更为简明。

我们首先回顾文献中研究过的一系列结构因果剃刀。每个定义都分为三部分： *[类]* 指定满足预期条件的有向无环图类， *[性质]* 描述 *[类]*, 中的每个有向无环图，而 *[剃刀]* 则说明真实因果模型（$\mathcal{G}^*$， $\mathcal{P}$）应满足的因果剃刀。谈及因果剃刀时，我们通常指的是 *[剃刀]* 部分。[5] 为方便读者查阅，因果剃刀的定义均置于方框中。

第一个候选原则是因果马尔可夫条件（CMC），它是因果发现最根本的支柱。

定义 3.1.（马尔可夫）对于 V 上的任意联合概率分布 $\mathcal{P}$,

- *[类]*$\mathrm{CMC}(\mathcal{P})$: $= \{\mathcal{G} \in \mathrm{DAG(v)}: \mathrm{I}(\mathcal{G}) \subseteq \mathrm{I}(\mathcal{P})\}$;
- *[性质]* 若有向无环图 $\mathcal{G}$ 满足 $\mathcal{P}$，则它是马尔可夫性的（相对于 $\mathcal{P}$）： $\mathcal{G} \in \mathrm{CMC}(\mathcal{P})$;
- *[剃刀]* $(\mathcal{G}^*, \mathcal{P})$ 满足因果马尔可夫条件，当且仅当 $\mathcal{G}^* \in \mathrm{CMC}(\mathcal{P})$。

CMC 要求，真实有向无环图（通过 d-分离）蕴含的条件独立关系是联合分布中成立的条件独立关系的子集。CMC 的定义源自全局马尔可夫条件（即 $\mathrm{I}(\mathcal{G}) \subseteq \mathrm{I}(\mathcal{P})$），后者等价于以下两个条件中的每一个。[6] 首先，若 $\mathcal{G}$ 满足相对于 $\mathcal{P}$的局部马尔可夫假设，则

$$X_i \perp\!\!\!\perp_{\mathcal{P}} \mathbf{X}_{\mathrm{nd}(i,\mathcal{G})} \setminus \mathbf{X}_{\mathrm{pa}(i,\mathcal{G})} \,|\, \mathbf{X}_{\mathrm{pa}(i,\mathcal{G})} \quad \text{for every } X_i \in \mathbf{V}. \tag{5}$$

换言之，在给定其所有父节点的条件下，每个变量都独立于其非后代节点。该条件有助于通过著名的马尔可夫分解重写联合概率分布 $\mathcal{P}$:

$$\mathcal{P}(\mathbf{V}) = \prod_{i \in \mathbf{v}} \mathcal{P}(X_i \,|\, \mathbf{X}_{\mathrm{pa}(i,\mathcal{G})}). \tag{6}$$

请注意，仅凭 CMC 无法达到识别真实马尔可夫等价类这一目标（除非是平凡情形，即 $\mathrm{I}(\mathcal{P}) = \varnothing$）。这是因为，如果 $\mathcal{G}$具有马尔可夫性，那么每个包含 $\mathcal{G}$ 的超图 $\mathcal{G}'$ （即 $\mathrm{E}(\mathcal{G}') \subseteq \mathrm{E}(\mathcal{G})$）也具有马尔可夫性。此外，每个完全有向无环图显然都具有马尔可夫性，因为它不会通过 d-分离蕴含任何条件独立关系。因此，$\mathrm{CMC}(\mathcal{P})$仍然留下一个巨大的假设空间供我们进行因果搜索。为了

[5]按照 [Zhang 2013], *[*关于术语的使用，因果剃刀的*]* 性质 *[*和*]* 剃刀 *[*可以分别解释为方法论假设和本体论假设。更准确地说，将因果剃刀（例如马尔可夫剃刀）视为一种方法论假设（或原则），意味着根据该假设，不满足*]* 性质 *[*的有向无环图（例如非马尔可夫图）属于应当被拒绝的假设。另一方面，将因果剃刀视为本体论假设，意味着假定真实的因果模型满足相关的*]* 性质[6]（例如因果马尔可夫条件）。 [三种马尔可夫条件等价性的证明可见于 2019]Lauritzen。

进一步缩小搜索空间，一种广受讨论的因果剃刀是 [Spirtes 等人 2000]提出的因果忠实性条件（CFC）。

定义 3.2。（忠实性）对于任意联合概率分布 $\mathcal{P}$,

- *[类]*CFC($\mathcal{P}$)：= {$\mathcal{G} \in$ CMC($\mathcal{P}$)：I($\mathcal{P}$) $\subseteq$ I($\mathcal{G}$)};
- *[性质]* 若一个 DAG $\mathcal{G}$ 对 $\mathcal{P}$是忠实的，则 $\mathcal{G} \in$ CFC($\mathcal{P}$);
- *[剃刀]* ($\mathcal{G}^*$,$\mathcal{P}$)满足因果忠实性条件，当且仅当 $\mathcal{G}^* \in$ CFC($\mathcal{P}$)。

通过假设 CMC 的逆命题，CFC 进一步要求 I($\mathcal{G}^*$) = I($\mathcal{P}$)。[7] 换言之，$\mathcal{P}$ 中成立的所有条件独立关系，都可以通过 d-分离由 $\mathcal{G}^*$ 完美解释。记 $\Psi_{\mathcal{G},\mathcal{P}}$ = I($\mathcal{P}$) \I($\mathcal{G}$) 为 $\mathcal{G} \in$ CMC($\mathcal{P}$)中不忠实的条件独立集合。有时我们写作 I($\mathcal{P}$) = I($\mathcal{G}$) $\cup$ $\Psi_{\mathcal{G},\mathcal{P}}$ 以突出 I($\mathcal{G}$) 与$\Psi_{\mathcal{G},\mathcal{P}}$之间的不交并。从方法论角度而言，当 $\mathcal{G} \in$ CFC($\mathcal{P}$)而 $\mathcal{G}'$ 具有非空的$\Psi_{\mathcal{G}',\mathcal{P}}$时，就解释力而言，$\mathcal{G}$ 严格优于 $\mathcal{G}'$，因此应当拒绝 $\mathcal{G}'$。

当 CFC 得到满足时，识别真实MEC 是一个可以实现的目标，因为 MEC($\mathcal{G}^*$) =CFC($\mathcal{P}$) 成立。值得庆幸的是，在大样本极限下，违反 CFC 的情况具有零勒贝格测度。[8] 然而，正如 Andersen [2013], 所论证的，非忠实性并不会以小概率消失，因为科学模型的某些参数值可能通过长期动态均衡获得，而这种均衡可能导致违反 CFC。此外，正如 Uhler 等人 [2013], 所表明的，通过假设检验从观察数据中学习条件独立关系容易出错，因此在有限样本中，几乎违反忠实性的情况十分常见。回到图1中的例子，$\mathcal{G}^* \notin$ 由于 C 与 T 之间几乎独立，而这种独立性并非由 $\mathcal{G}^*$ 尽管 d-分离所蕴含，因此 CFC($\mathcal{P}$)。

在维持同一目标的同时承认 CFC 的频繁近似违反后，研究人员受到促使，转而采用严格弱于 CFC 的因果剃刀。最早的方法之一是 Ramsey 等人 [2006], 提出的受限忠实性假设，该假设可以分解为邻接关系忠实性和方向忠实性。[9]

De义 3.3.（邻接忠实性）对于 V 上的任意联合概率分布 $\mathcal{P}$,

- *[性质]* 若有向无环图 $\mathcal{G}$ 对 $\mathcal{P}$ 邻接忠实，则对于每个 (i, j) $\in$ sk($\mathcal{G}$)，以及每个 s $\subseteq$ v \ {i, j}，都有 $\langle$Xi, X j ∣Xs$\rangle \notin$ I($\mathcal{P}$) ;
- *[类]* AdjF($\mathcal{P}$)：= {$\mathcal{G} \in$ CMC($\mathcal{P}$)： $\mathcal{G}$ 对 $\mathcal{P}$} 邻接忠实;
- *[剃刀]* ($\mathcal{G}^*$,$\mathcal{P}$) 满足因果邻接忠实性假设，当且仅当 $\mathcal{G}^* \in$ AdjF($\mathcal{P}$)。

[7]不采用常见做法将 CFC 定义为 CMC 的逆命题（或 CFC($\mathcal{P}$)：= {$\mathcal{G} \in$ DAG(v)：I($\mathcal{P}$) $\subseteq$I($\mathcal{G}$)}），而是将 CFC 定义为 CMC 与其逆命题的合取（或 CFC($\mathcal{P}$)：= {$\mathcal{G} \in$ CMC(v)：I($\mathcal{P}$) $\subseteq$ I($\mathcal{G}$)}，如定义 3.2 所示），这样有助于简化论述，因为不存在忠实但非马尔可夫性的有向无环图。下文定义的每一种因果剃刀的 *[类]* 也将采用同样的约定。[8]参见 Spirtes 等人, [2000, 第 41-2] 页。[9]遵循 CMC 和 CFC 的定义，我们将每种 *[剃刀]* 命名为以“因果”一词开头，以表明满足相关 *[性质]*的是真正的因果模型。不过，为简洁起见，后文通常会省略这一命名惯例。

De义 3.4.（方向忠实性）对于 V 上的任意联合概率分布 $\mathcal{P}$ ，

- *[性质]* 如果有向无环图 $\mathcal{G}$ （相对于 $\mathcal{P}$）满足定向忠实性，则对于 $\mathcal{G}$ 中的每个非屏蔽三元组 (i, j, k)，(a) 如果 $\mathcal{G}$ 中成立 $i \to j \leftarrow k$，那么对于每个满足 $s \subseteq v \setminus \{i, k\}$ 且 $j \in s$ 的 s，$\langle Xi, Xk \mid X_s \rangle \notin I(\mathcal{P})$；(b) 否则，对于每个满足 $s \subseteq v \setminus \{i, k\}$ 且 $j \notin s$ 的 s，$\langle Xi, Xk \mid X_s \rangle \notin I(\mathcal{P})$；

- *[类]* $OriF(\mathcal{P}) := \{\mathcal{G} \in CMC(\mathcal{P}) : \mathcal{G}$ 相对于 $\mathcal{P}\}$满足定向忠实性；
- *[剃刀]* $(\mathcal{G}^*, \mathcal{P})$ 满足因果定向忠实性假设，当且仅当 $\mathcal{G}^* \in OriF(\mathcal{P})$。

定义 3.5。（受限忠实性）对于任意联合概率分布 $\mathcal{P}$,

- *[类]* $ResF(\mathcal{P}) := AdjF(\mathcal{P}) \cap OriF(\mathcal{P})$;
- *[性质]* 若满足 $\mathcal{G} \in ResF(\mathcal{P})$，则有向无环图 $\mathcal{G}$ 是（相对于 $\mathcal{P}$）限制忠实的；

若满足 $\mathcal{G}^* \in ResF(\mathcal{P})$，则• *[剃刀]* $(\mathcal{G}^*, \mathcal{P})$ 满足因果限制忠实性假设。

根据 Verma 和 Pearl 的经典结果，对于任意给定的有向无环图 $\mathcal{G}$ ，其 [1990], 马尔可夫等价类$\mathcal{G}$可以通过其骨架 $sk(\mathcal{G})$ 和其中无屏蔽碰撞点的集合唯一确定。请注意，限制忠实性假设保证 $I(\mathcal{P})$ 能够正确识别 $\mathcal{G}^*$ 中的邻接关系和无屏蔽碰撞点（分别通过邻接忠实性和原始忠实性）。因此，在关系忠实性假设下，$\mathcal{G}^*$马尔可夫等价类$\mathcal{G}^*$可以被识别。正如 Ramsey 等人 [2006], 所证明的，在关系忠实性成立时，著名的 PC算法（由 Spirtes 和 Glymour[1991]提出）能够正确识别$\mathcal{G}^*$马尔可夫等价类。对于图 1 中我们先前的五节点示例， $\mathcal{G}^*$ 并非邻接忠实，因为 C 与 T 之间的邻接关系对应着一种（几乎）不忠实的独立性。因此， $\mathcal{G}^*$ 也不是限制忠实的。

识别 $MEC(\mathcal{G}^*)$ 的另一种途径是 Raskutti 和 Uhler [2018] 提出的最稀疏马尔可夫表示（SMR）假设。该假设根据 $\mathcal{G}^*$ 的稀疏性来定义，或者等价地说，根据 [的简约性来定义，见 Forster 等人， 2020]。

定义 3.6。（简约性）对于任意联合概率分布 $\mathcal{P}$,

- *[类]* $Fr(\mathcal{P})$: $= \{\mathcal{G} \in CMC(\mathcal{P}) : \neg\exists \mathcal{G}' \in CMC(\mathcal{P})$，使得 $|E(\mathcal{G}')| < |E(\mathcal{G})|\}$;
- *[性质]* 若有向无环图 $\mathcal{G}$ 是简约的（相对于 $\mathcal{P}$），则 $\mathcal{G} \in Fr(\mathcal{P})$;
- *[剃刀]* $(\mathcal{G}^*, \mathcal{P})$ 满足因果简约假设，当且仅当 $\mathcal{G}^* \in Fr(\mathcal{P})$。

定义 3.7.（唯一简约性 / 最稀疏马尔可夫表示）对于任意联合概率分布 $\mathcal{P}$,

- *[类]* $\mathrm{uFr}(\mathcal{P}) := \begin{cases} \mathrm{Fr}(\mathcal{P}), & \text{若其中所有有向无环图都属于同一个马尔可夫等价类，则} \\ \varnothing & \text{否则;} \end{cases}$
- *[性质]* 若某个有向无环图 $\mathcal{G}$ 满足 $\mathcal{G} \in \mathrm{uFr}(\mathcal{P})$，则称其为（相对于 $\mathcal{P}$）u-简约的;
- *[若* $\mathcal{G}^* \in \mathrm{uFr}(\mathcal{P})$，则剃刀*]* $(\mathcal{G}^*, \mathcal{P})$满足因果唯一简约性假设。

从本体论的角度看，简约性假设 $\mathcal{G}^*$ 是最稀疏的马尔可夫有向无环图之一。从方法论上说，它指出只有最稀疏的马尔可夫有向无环图才是可接受的假设。回到图1和图2中的两个有向无环图，$\mathcal{G}^*$ 是简约有向无环图，而 $\mathcal{G}'$ 不是。[10] 另一方面，u-简约性不仅要求简约性，还要求马尔可夫等价类的唯一性（即，所有简约有向无环图都马尔可夫等价）。因此，根据定义，u-简约性强于其非唯一版本。相应地，u-简约性要求 $\mathrm{MEC}(\mathcal{G}^*)$ 与 $\mathrm{Fr}(\mathcal{P})$ 等价，因此在正确假设 u-简约性的情况下，$\mathrm{MEC}(\mathcal{G}^*)$ 的识别是显而易见的。实践中，Raskutti 和 Uhler2018] 提出的最稀疏排列（SP）算法通过枚举所有可能的排列并找出由此导出的最稀疏有向无环图，返回 $\mathrm{Fr}(\mathcal{P})$。因此，当满足 u-简约性时，SP 算法的正确性显而易见。[11]

现在我们转而讨论文献中研究的其他一些因果剃刀。其中一个候选条件是极小性条件，有时也称为极小I映射（例如，Verma 和 Pearl [1988]）。我们沿用 Zhang [2013]的做法，并按照 [Spirtes 等人 2000]的称法，将这一因果剃刀称为 SGS-极小性。

定义 3.8。（SGS-极小性）对于任意联合概率分布 $\mathcal{P}$,

- *[类]* $\mathrm{SGSM}(\mathcal{P}) := \{\mathcal{G} \in \mathrm{CMC}(\mathcal{P}) : \neg\exists \mathcal{G}' \in \mathrm{CMC}(\mathcal{P})$ 使得 $\mathrm{E}(\mathcal{G}') \subset \mathrm{E}(\mathcal{G})\}$;
- *[性质]* 一个有向无环图 $\mathcal{G}$ 相对于 $\mathcal{P}$ 是 SGS-极小的，当且仅当 $\mathcal{G} \in \mathrm{SGSM}(\mathcal{P})$;
- *[剃刀]* $(\mathcal{G}^*, \mathcal{P})$ 满足因果 SGS-极小性假设，当且仅当 $\mathcal{G}^* \in \mathrm{SGSM}(\mathcal{P})$。

通俗地说，SGS-极小性要求不存在任何马尔可夫性子图。与简约性类似，SGS-极小性是一种边极小性条件，但其定义依据是集合包含，而非基数。这一因果剃刀在基于排列的因果搜索算法（例如 SP）中发挥着关键作用。正如 Verma 和 Pearl [1988], 所证明的，由 v 的拓扑排序和一个半拟阵（例如联合概率分布）构造出的有向无环图必定是 SGS-极小的。[12] 此外，正如 Zhang 和 Spirtes [2011],所论证的，SGS-极小性是一个相当稳妥的假设，尤其是在因果关系的干预主义解释下。[13]

另一种备受关注的因果剃刀，是 Pearl [2009] 研究的极小性原则。按照 [Zhang 的命名惯例，2013], 我们将这种因果剃刀称为 Pearl极小性，简称 P-极小性。

[10]详情请见例 4.4 中关于 $\mathrm{I}(\mathcal{P})$ 的说明，以验证不存在比 $\mathcal{G}^*$ 更稀疏的马尔可夫性有向无环图。[11]具体而言，假设 $\mathcal{P}$ 是一个拟阵（见附录 A），SP 会枚举 v 上的所有排列，其中每种排列都会导出一个 SGS 极小有向无环图。[12]半拟阵的定义见附录 A。[13]读者或许会疑惑，为何上文没有介绍 SGS 极小性的唯一变体。关于它为何是一种无趣的因果剃刀，请见脚注 23。

定义 3.9。（P-极小性）对于任意联合概率分布 $\mathcal{P}$,

- *[类]* $\mathrm{PM}(\mathcal{P}) := \{\mathcal{G} \in \mathrm{CMC}(\mathcal{P}) : \neg\exists \mathcal{G}' \in \mathrm{CMC}(\mathcal{P})$ 使得 $\mathrm{I}(\mathcal{G}) \subset \mathrm{I}(\mathcal{G}')\}$;
- *[性质]* 若有向无环图 $\mathcal{G}$ 满足 P-极小性（相对于 $\mathcal{P}$），则称其为 P-极小的; $\mathcal{G} \in \mathrm{PM}(\mathcal{P})$;
- *[剃刀]* $(\mathcal{G}^*, \mathcal{P})$ 满足因果 P-极小性假设，当且仅当 $\mathcal{G}^* \in \mathrm{PM}(\mathcal{P})$。

鉴于马尔可夫有向无环图是一种旨在解释 $\mathrm{I}(\mathcal{P})$ 中条件独立关系的假设，P-极小性要求由 $\mathcal{G}$ 未能解释的条件独立关系集合（即 $\Psi_{\mathcal{G},\ \mathcal{P}}$）在集合论意义上是极小的。从方法论角度而言，如果一个马尔可夫有向无环图 $\mathcal{G}$ 相较于另一个马尔可夫有向无环图 $\mathcal{G}'$ 能解释更多的条件独立关系（即 $\mathrm{I}(\mathcal{G}') \subset \mathrm{I}(\mathcal{G}) \subseteq \mathrm{I}(\mathcal{P})$），则应当拒绝后者。Zhang [2013] 提出了若干论据，支持将 P-极小性视为一种本体论假设，尤其是在 $\mathcal{G}^*$ 相对稀疏、三角形较少时。此外，Lin和Zhang [2020] 通过采用学习理论方法表明，违反 P-极小性必然会牺牲某些重要的收敛性质。

接下来，类似于简约性与u-简约性之间的关系，我们引入带有唯一性条件的强化形式P-极小性。

定义3.10。（唯一P-极小性）对于任意联合概率分布 $\mathcal{P}$,

- *[类]* $\mathrm{uPM}(\mathcal{P}) := \begin{cases} \mathrm{PM}(\mathcal{P}), & \text{当且仅当PM}(\mathcal{P})\text{中的所有有向无环图都属于同一个马尔可夫等价类;} \\ \varnothing & \text{否则;} \end{cases}$
- *[性质]* 若有向无环图 $\mathcal{G}$ 满足 $\mathcal{G} \in \mathrm{uPM}(\mathcal{P})$，则称其为（相对于 $\mathcal{P}$）u-P-极小;
- *[若* $\mathcal{G}^* \in \mathrm{uPM}(\mathcal{P})$，则称剃刀*]* $(\mathcal{G}^*, \mathcal{P})$满足因果u-P-极小性假设。

与讨论过的其他因果剃刀不同，除了Lin和Zhang[2020]以及Lam等人 [2022]之外，u-P-极小性并未受到太多关注。尽管不太受关注，后两位作者证明了u-P-极小性与CFC在逻辑上等价。我们将在附录B中重新阐述他们简洁的证明，并在第6节讨论这种逻辑等价性带来的算法层面影响。

我们现在转向一种因果剃刀，它已在 [Spirtes和Zhang、 2015], [Zhang和Spirtes、2008], 以及 [Zhang 的研究中得到广泛探讨， 2013], 即三角忠实性假设。

定义 3.11。（三角忠实性）给定一个关于 V $\mathcal{P}$的联合概率分布 ,

- *[性质]* 一个有向无环图 $\mathcal{G}$ 是（三角）忠实的（相对于 $\mathcal{P}$），当且仅当对于 $\mathcal{G}$中的每个有屏蔽三元组 (i, j, k)，(a) 若 $i \to j \leftarrow k$ 在 $\mathcal{G}$中成立，则对于任意满足 $s \subseteq V \setminus \{i, k\}$ 且 $j \in s$ 的 s，$\langle X_i, X_k \mid X_s \rangle \notin \mathrm{I}(\mathcal{P})$；(b) 否则，对于任意满足 $s \subseteq V \setminus \{i, k\}$ 且 $j \notin s$ 的 s，$\langle X_i, X_k \mid X_s \rangle \notin \mathrm{I}(\mathcal{P})$;
- *[类]* $\mathrm{TriF}(\mathcal{P}) := \{\mathcal{G} \in \mathrm{CMC}(\mathcal{P}) : \mathcal{G}$ 相对于 $\mathcal{P}\}$是三角忠实的;
- *[剃刀]* $(\mathcal{G}^*, \mathcal{P})$ 满足因果三角忠实性假设，当且仅当 $\mathcal{G}^* \in \mathrm{TriF}(\mathcal{P})$。

读者可能会发现，三重忠实性与原始忠实性非常相似，前者定义于三角形之上，后者定义于无屏蔽三元组之上。尽管如此，三重忠实性是邻接忠实性的逻辑推论（但不是原始忠实性的逻辑推论）。关于三重忠实性，有两点值得注意。第一，它与非忠实性的可检测性有关。确切地说，任何一种 *[剃刀]* 的违反都可以分为可检测和不可检测两类：若其对应的 *[类]* 为空，则为可检测；若 *[类]* 非空但不包含 $\mathcal{G}^*$，则为不可检测。正如 Zhang 和 Spirtes [2008], 所证明的，任何对 CFC 的不可检测违反都必定是违反三重忠实性的结果。另一方面，正如[Zhang 2013], 所示，SGS-极小性与三重忠实性的合取蕴含 P-极小性。这支持了他的主张：当 $\mathcal{G}^*$ 相对稀疏且三角形数量较少时，假设其满足 P-极小性相对安全。

本节定义的十一个因果剃刀并未涵盖文献中提出的所有因果剃刀。例如， [Zhalama 等人 2019] 提出的 NOI-极小性要求，真实 DAG 在所有马尔可夫有向无环图中蕴含最多的条件独立关系。 [Ng 等人2021]提出的单一有屏蔽／无屏蔽碰撞点忠实性，弱化了三元／原始忠实性假设。建议读者将文献中的其他候选方案与上述定义进行关联，以更全面地阐释其逻辑。我们之所以重点讨论上述十一个因果剃刀，主要是因为它们具有紧密的逻辑层级，我们将在下一节探讨这些层级。

4 结构因果剃刀的逻辑层级

本节将从逻辑强度方面比较上述十一个因果剃刀。具体而言，可以通过考察它们的 *[类]*之间可能存在的成对包含关系来进行比较。两个 *[类]*之间存在包含关系的情形，大多可以通过简单的证明得出（有些甚至由定义直接成立）。除了附录B中的一项较新的结果——CFC 与 u-P-最小性在逻辑上等价——我们不再重述具体证明。相反，对于每一种不满足包含关系的情形，我们都会给出一个反例。

Th定理 4.1。对于任意联合概率分布 $\mathcal{P}$，以下陈述成立：

(a) $\mathrm{CFC}(\mathcal{P}) \subseteq \mathrm{ResF}(\mathcal{P}) = \mathrm{AdjF}(\mathcal{P}) \cap \mathrm{OriF}(\mathcal{P}) \subseteq \mathrm{uFr}(\mathcal{P}) \subseteq \mathrm{Fr}(\mathcal{P}) \subseteq \mathrm{PM}(\mathcal{P}) \subseteq \mathrm{SGSM}(\mathcal{P}) \subseteq \mathrm{CMC}(\mathcal{P})$；(b) $\mathrm{AdjF}(\mathcal{P}) \subseteq \mathrm{TriF}(\mathcal{P})$；(c) $\mathrm{AdjF}(\mathcal{P}) \subseteq \mathrm{Fr}(\mathcal{P})$。

证明。对于 (a)，定义 3.5、3.7、3.8 分别给出 $\mathrm{ResF}(\mathcal{P}) = \mathrm{AdjF}(\mathcal{P}) \cap \mathrm{OriF}(\mathcal{P})$、$\mathrm{uFr}(\mathcal{P}) \subseteq \mathrm{Fr}(\mathcal{P})$ 以及 $\mathrm{SGSM}(\mathcal{P}) \subseteq \mathrm{CMC}(\mathcal{P})$。$\mathrm{CFC}(\mathcal{P}) \subseteq \mathrm{ResF}(\mathcal{P})$ 已由 [Ramsey 等人，2006], $\mathrm{ResF}(\mathcal{P}) \subseteq \mathrm{uFr}(\mathcal{P})$ 已由 [Raskutti 和 Uhler，2018], $\mathrm{Fr}(\mathcal{P}) \subseteq \mathrm{PM}(\mathcal{P})$ 已由 [Forster 等人，2020], 以及 $\mathrm{PM}(\mathcal{P}) \subseteq \mathrm{SGSM}(\mathcal{P})$ 已由[Zhang, 2013]证明。(b) 根据定义 3.11 成立。

对于 (c)，用归谬法，假设 $\mathcal{G} \in \mathrm{DAG(v)}$ 且 $\mathcal{G} \in \mathrm{AdjF}(\mathcal{P}) \setminus \mathrm{Fr}(\mathcal{P})$。因此，存在 $\mathcal{G}' \in \mathrm{CMC}(\mathcal{P})$，使得 $|\mathrm{E}(\mathcal{G})| > |\mathrm{E}(\mathcal{G}')|$。必定存在 $(i, j) \in \mathrm{sk}(\mathcal{G})$，使得 $(i, j) \notin \mathrm{sk}(\mathcal{G}')$。鉴于 $(i, j) \notin \mathrm{sk}(\mathcal{G}')$，$\langle X_i, X_j \mid X_{\mathbf{s}} \rangle \in \mathrm{I}(\mathcal{G}') \subseteq \mathrm{I}(\mathcal{P})$，其中某个 $\mathbf{s} \subseteq \mathrm{v} \setminus \{i, j\}$。然而，由于 $(i, j) \in \mathrm{sk}(\mathcal{G})$，这与$\mathcal{G} \in \mathrm{AdjF}(\mathcal{P})$矛盾。□

Theo注 4.2. Lam 等人 *[2022]* 对于任意联合概率分布 $\mathcal{P}$，$\mathrm{CFC}(\mathcal{P}) = \mathrm{uPM}(\mathcal{P}$).

证明。见附录B。 □

接下来，我们转向子集关系失效的反例。为简化论述，下文仅列出独立性模型中两个单元素集合之间成立的条件独立关系（例如，$\langle X, Y \mid Z\rangle$）。严谨的读者可以验证，以下每个示例的独立性模型 $\mathrm{I}(\mathcal{P})$ 都会诱导出一个（组合的）拟阵（见附录 A）。

例 4.3. 存在一个联合概率分布 $\mathcal{P}$，使得 $\mathrm{ResF}(\mathcal{P}) \setminus \mathrm{CFC}(\mathcal{P}) \neq \varnothing$。

考虑图 3 中的有向无环图 $\mathcal{G}^*$。令 $\mathcal{P}$ 为联合概率分布，其中 $\mathrm{I}(\mathcal{P}) = \mathrm{I}(\mathcal{G}^*) \cup \{\langle X_1, X_6\rangle\}$。显然，$\mathcal{G}^* \notin \mathrm{CFC}(\mathcal{P})$。但是，$X_1 \perp\!\!\!\perp_{\mathcal{P}} X_6$ 与 $\mathcal{G}^*$中的任何邻接关系或非屏蔽三元组都无关。因此，$\mathcal{G}^* \in \mathrm{ResF}(\mathcal{P})$。[14] □

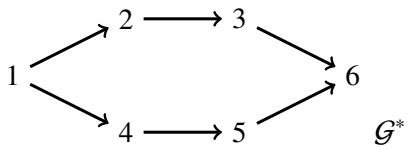


图 3：由于 $\Psi_{\mathcal{G}^*,\,\mathcal{P}} = \{\langle X_1, X_6\rangle\}$，$\mathcal{G}^* \in \mathrm{ResF}(\mathcal{P}) \setminus \mathrm{CFC}(\mathcal{P})$

例 4.4. 存在一个联合概率分布 $\mathcal{P}$，使得 $\mathrm{uFr}(\mathcal{P}) \setminus \mathrm{TriF}(\mathcal{P}) \neq \varnothing$。

此例引自 [Lam 等人，2022年，补充材料]。[15] 考虑图 4 中的有向无环图 $\mathcal{G}^*$。令$\mathcal{P}$ 为满足 $\mathrm{I}(\mathcal{P}) = \mathrm{I}(\mathcal{G}^*) \cup \Psi_{\mathcal{G}^*,\,\mathcal{P}}$ 的联合概率分布，其中

$$\Psi_{\mathcal{G}^*,\mathcal{P}} = \begin{Bmatrix} \psi_1 : \langle X_1, X_5\rangle, & \psi_2 : \langle X_1, X_2 \mid \{X_5\}\rangle, \\ \psi_3 : \langle X_1, X_3 \mid \{X_5\}\rangle, & \psi_4 : \langle X_1, X_5 \mid \{X_2\}\rangle, \\ \psi_5 : \langle X_1, X_5 \mid \{X_3\}\rangle, & \psi_6 : \langle X_1, X_2 \mid \{X_3, X_5\}\rangle, \\ \psi_7 : \langle X_1, X_3 \mid \{X_2, X_5\}\rangle, & \psi_8 : \langle X_1, X_5 \mid \{X_2, X_3\}\rangle \end{Bmatrix}.$$

读者可以验证，每个马尔可夫性有向无环图要么与 $\mathcal{G}^*$ 马尔可夫等价，要么比 $\mathcal{G}^*$ 更稠密。[16] 因此，$\mathcal{G}^* \in \mathrm{uFr}(\mathcal{P})$。接下来，由于三角形 (1, 4, 5) 以及 ψ_1，有 $\mathcal{G}^* \notin \mathrm{TriF}(\mathcal{P})$。事实上，还可以证明 $\mathrm{TriF}(\mathcal{P}) = \varnothing$。[17] □

例 4.5。存在一个联合概率分布 $\mathcal{P}$，使得

(a) $\mathbb{F}\mathbb{r}(\mathcal{P}) \setminus \cup\mathbb{F}\mathbb{r}(\mathcal{P}) \neq \varnothing$;

（b）$\mathrm{TriF}(\mathcal{P}) \setminus \mathrm{uFr}(\mathcal{P}) \neq \varnothing$;

[14]这个例子可以通过路径抵消，轻松构造为一个线性高斯因果模型。鉴于 $\boldsymbol{\theta}_{\mathcal{G}^*} = \{\theta_{12}, \theta_{14}, \theta_{23}, \theta_{36}, \theta_{45}, \theta_{56}\}$ $\theta_{12}\theta_{23}\theta_{36} = -\theta_{14}\theta_{45}\theta_{56}$ $1 \to 2 \to 3 \to 6$ $1 \to 4 \to 5 \to 6$，我们设定 ( )。因此，$\mathcal{G}^*$ 中的两条有向路径相互抵消，并导出 $\langle X_1, X_6\rangle \in \Psi_{\mathcal{G}^*,\,\mathcal{P}}$。[15]要将此例构造为线性高斯因果模型（$\mathcal{G}^*$，$\mathcal{P}$），设定 $-\theta_{15} = \theta_{14}\theta_{45}$，使得两条有向路径 $1 \to 5$ 和 $1 \to 4 \to 5$ 相互抵消，并导出不忠实的 $X_1 \perp\!\!\!\perp_{\mathcal{P}} X_5$。其他不忠实的条件独立关系（即 $\{\psi_2, \ldots, \psi_8\}$）可由 $\mathrm{I}(\mathcal{G}^*) \cup \{\langle X_1, X_5\rangle\}$ 结合拟阵公理（包括交集和组合）推出。[16]验证这一点的一种简单方法是，按照 [Verma 和 Pearl, 1990] 或 [Raskutti 和 Uhler, 2018] 中讨论的有向无环图构造方法，从顶点集合 v 上的每个排列 π 导出一个有向无环图 $\mathcal{G}_\pi$。[17]通过枚举所有 SGS 极小有向无环图（使用脚注 16 中讨论的方法），可以验证它们都不是三角形忠实的。接下来注意，如果 $\mathcal{G} \notin \mathrm{TriF}(\mathcal{P})$，那么对于任意满足 $\mathrm{E}(\mathcal{G}) \subseteq \mathrm{E}(\mathcal{G}')$ 的 $\mathcal{G}' \in \mathrm{DAG}(\mathrm{v})$，都有 $\mathcal{G}' \notin \mathrm{TriF}(\mathcal{P})$。由于每个马尔可夫性有向无环图都是某个 SGS 极小有向无环图的超图，因此没有马尔可夫性有向无环图是三角形忠实的。因此，$\mathrm{TriF}(\mathcal{P}) = \varnothing$。

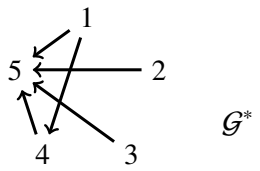


图 4： $\mathcal{G}^* \in \text{uFr}(\mathcal{P}) \setminus \text{TriF}(\mathcal{P})$，这是由于 $\langle X_1, X_5 \rangle \in \Psi_{\mathcal{G}^*, \mathcal{P}}$。

(c) $\text{OriF}(\mathcal{P}) \setminus \text{uFr}(\mathcal{P}) \neq \emptyset$;
(d) $\text{TriF}(\mathcal{P}) \setminus \text{AdjF}(\mathcal{P}) \neq \emptyset$;
(e) $\text{OriF}(\mathcal{P}) \setminus \text{TriF}(\mathcal{P}) \neq \emptyset$.

该例来自 Forster 等人 [2020]。[18] 考虑图 5 中的 $\mathcal{G}^*$ 和 $\mathcal{G}'$。令 $\mathcal{P}$ 为联合概率分布，其中 $\text{I}(\mathcal{P}) = \text{I}(\mathcal{G}^*) \cup \{\langle X_1, X_4 \rangle\}$。读者可以轻松验证，$\mathcal{G}^*, \mathcal{G}' \in \text{Fr}(\mathcal{P})$。[19] 然而，$\text{uFr}(\mathcal{P}) = \emptyset$，因为 $\mathcal{G}' \notin \text{MEC}(\mathcal{G}^*)$。因此，(a) 得证。此外，我们有 $\mathcal{G}^* \in \text{OriF}(\mathcal{P}) \cap \text{TriF}(\mathcal{P})$，因为 $\mathcal{G}^*$ 的唯一不忠实条件独立是 $X_1 \perp\!\!\!\perp_{\mathcal{P}} X_4$，它既不与 $\mathcal{G}^*$中的任何非屏蔽三元组相关，也不与任何三角形相关。然而，由于 $(1, 4) \in \text{sk}(\mathcal{G}^*)$ 以及该不忠实条件独立，$\mathcal{G}^* \notin \text{AdjF}(\mathcal{P})$。因此，(b)-(d) 也得证。最后，$\mathcal{G}'$ 的唯一不忠实条件独立 $X_1 \perp\!\!\!\perp_{\mathcal{P}} X_3 \mid \{X_2\}$ 不与任何非屏蔽三元组相关，但它在三角形 (1, 3, 2) 处构成了对三重忠实性的违反。因此，(e) 立即成立。□



图 5： $\mathcal{G}^*, \mathcal{G}' \in \text{Fr}(\mathcal{P})$，但 $\text{uFr}(\mathcal{P}) = \emptyset$。

例 4.6。存在一个联合概率分布 $\mathcal{P}$，使得

(a) $\mathbb{uFr}(\mathcal{P}) \setminus \mathbb{OriF}(\mathcal{P}) \neq \emptyset$;

(b) $\mathbb{AdjF}(\mathcal{P}) \setminus \mathbb{OriF}(\mathcal{P}) \neq \emptyset$;

(c) $\mathbb{PM}(\mathcal{P}) \setminus \mathbb{Fr}(\mathcal{P}) \neq \emptyset$.

考虑图 6 中的 $\mathcal{G}^*$。令 $\mathcal{P}$ 为联合概率分布，其中 $\text{I}(\mathcal{P}) = \text{I}(\mathcal{G}^*) \cup \{\langle X_1, X_4 \rangle\}$。[20] 留给读者验证 $\mathcal{G}^* \in \text{uFr}(\mathcal{P})$。[21] 接下来请注意，$\mathcal{G}^*$ 唯一的不忠实条件独立 $X_1 \perp\!\!\!\perp_{\mathcal{P}} X_4$ 与 $\mathcal{G}^*$ 中的任何邻接关系都无关。因此，如 (a) 所示，我们有 $\mathcal{G}^* \in \text{AdjF}(\mathcal{P})$。然而，这违反了原始忠实性，因为原始忠实性要求 X_1 和 X_4 与 X_2 和/或 X_3 条件相关。因此，如 (b) 所示，$\mathcal{G}^* \notin \text{OriF}(\mathcal{P})$。反过来，考虑图 6 中的 $\mathcal{G}'$，其中 $\text{I}(\mathcal{G}') = \{\langle X_1, X_4 \rangle\} \subset \text{I}(\mathcal{P})$。显然，$\mathcal{G}' \notin \text{Fr}(\mathcal{P})$，因为 $|\text{E}(\mathcal{G}^*)| < |\text{E}(\mathcal{G}')|$。此外，无法扩展 $\text{I}(\mathcal{G}')$ 以得到马尔可夫有向无环图。因此，如 (c) 所示，$\mathcal{G}' \in \text{PM}(\mathcal{P}) \setminus \text{Fr}(\mathcal{P})$。□

[18]为了将该例构造为线性高斯模型，设定 $-\theta_{14} = \theta_{12}\theta_{23}\theta_{34}$，使得 $1 \to 2 \to 3 \to 4$ 和 $1 \to 4$ 相互抵消，并诱导出不忠实的 $X_1 \perp\!\!\!\perp_{\mathcal{P}} X_4$。[19]$\text{Fr}(\mathcal{P})$ 可通过 SP 算法获得，该算法会枚举所有置换诱导的有向无环图，如脚注 16 所述。[20]与上文类似，在该线性高斯模型 $(\mathcal{G}^*, \mathcal{P})$ 中设定 $\theta_{12}\theta_{24} = -(\theta_{13}\theta_{34})$，两条路径 $1 \to 2 \to 4$ 和 $1 \to 3 \to 4$ 会相互抵消，并诱导出不忠实的 $X_1 \perp\!\!\!\perp_{\mathcal{P}} X_4$。[21]类似于脚注 19，我们通过 SP 算法获得 $\text{Fr}(\mathcal{P})$，并验证 $\text{Fr}(\mathcal{P})$ 中的所有有向无环图都属于 $\text{MEC}(\mathcal{G}^*)$。



图 6：$\Psi_{\mathcal{G}^*,\ \mathcal{P}} = \{\langle X_1, X_4\rangle\}$ 而 $\mathcal{G}' \in \mathrm{PM}(\mathcal{P}) \setminus \mathrm{Fr}(\mathcal{P})$。

例 4.7。存在一个联合概率分布 $\mathcal{P}$，使得 $\mathrm{AdjF}(\mathcal{P}) \setminus \mathrm{uFr}(\mathcal{P}) \neq \varnothing$。

令 $\mathcal{P}$ 为联合概率分布，其中 $\mathrm{I}(\mathcal{P}) = \{\langle X_1, X_3\rangle,\ \langle X_1, X_3 \mid \{X_2\}\rangle\}$。[22] 考虑图 7 中的两个马尔可夫有向无环图 $\mathcal{G}_0$、$\mathcal{G}_1$。由于它们各自的不忠实条件独立均与邻接关系无关，因此我们有 $\mathcal{G}_0$、$\mathcal{G}_1 \in \mathrm{AdjF}(\mathcal{P})$。然而，它们是简约有向无环图，且不属于同一个马尔可夫等价类。因此，$\mathrm{uFr}(\mathcal{P}) = \varnothing$。□



图 7：尽管 gh 两者都满足 $\mathcal{G}_0, \mathcal{G}_1 \in \mathrm{AdjF}(\mathcal{P})$，但它们不属于同一个马尔可夫等价类，因此 $\mathrm{uFr}(\mathcal{P}) = \varnothing$.

例 4.8。存在一个联合概率分布 $\mathcal{P}$，使得

(a) $\mathbb{CMC}(\mathcal{P}) \setminus \mathbb{SGSM}(\mathcal{P}) \neq \varnothing$;

(b) $\mathbb{SGSM}(\mathcal{P}) \setminus \mathbb{PM}(\mathcal{P}) \neq \varnothing$.

考虑图 8 中的三个有向无环图。令 $\mathcal{P}$ 为联合概率分布，其中 $\mathrm{I}(\mathcal{P}) = \{\langle X_1, X_3\rangle\}$。因此，$\mathcal{G}^* \in \mathrm{CFC}(\mathcal{P})$。注意，$\mathcal{G}' \in \mathrm{CMC}(\mathcal{P})$ 是 $\mathcal{G}^*$ 的超图。这意味着 $\mathcal{G}' \notin \mathrm{SGSM}(\mathcal{P})$。另一方面，$\mathcal{G}'' \in \mathrm{SGSM}(\mathcal{P})$，因为 $\mathcal{G}''$ 的任何子图都不具有马尔可夫性。然而，$\mathcal{G}'' \notin \mathrm{PM}(\mathcal{P})$，因为$\varnothing = \mathrm{I}(\mathcal{G}'') \subset \mathrm{I}(\mathcal{G}^*) = \mathrm{I}(\mathcal{P})$。
[23]
□

例 4.9。存在一个联合概率分布 $\mathcal{P}$，使得

(a) $\mathbb{OriF}(\mathcal{P}) \setminus \mathbb{SGSM}(\mathcal{P}) \neq \varnothing$;

(b) $\mathbb{TriF}(\mathcal{P}) \setminus \mathbb{SGSM}(\mathcal{P}) \neq \varnothing$.

[22]多项因果模型（$\mathcal{G}_0$，$\mathcal{P}$）的构造见附录 E.1。[23]忠实的（$\mathcal{G}^*$，$\mathcal{P}$）可以很容易地构造为线性高斯模型或多项因果模型。另一方面，这个例子告诉我们，$\mathcal{G}^*$ 和 $\mathcal{G}''$ 相对于 $\mathcal{P}$ 都是 SGS-极小的，但它们不属于同一个马尔可夫等价类。这意味着，即使满足 CFC，SGS 极小性的唯一变体仍然不成立。因此，唯一 SGS 极小性是一种过强且无趣的因果剃刀。

此外，请注意，$\mathcal{G}' \notin \mathrm{SGSM}(\mathcal{P})$，但 $\mathcal{G}'' \in \mathrm{SGSM}(\mathcal{P})$. 这表明 SGS-极小性并非 MEC 保持 .

更准确地说，如果同一 MEC 中的每个有向无环图都属于因果剃刀所指定的 *[类]*，或者都不属于该 *[类]*，那么该因果剃刀就是 MEC 保持的。类似地，这个例子也表明，三重忠实性并非 MEC 保持：$\mathcal{G}' \in \mathrm{TriF}(\mathcal{P})$，而 $\mathcal{G}'' \notin \mathrm{TriF}(\mathcal{P})$。读者可以验证，本文讨论的所有因果剃刀中，除了 SGS-极小性和三重忠实性之外，其余都是 MEC 保持的。

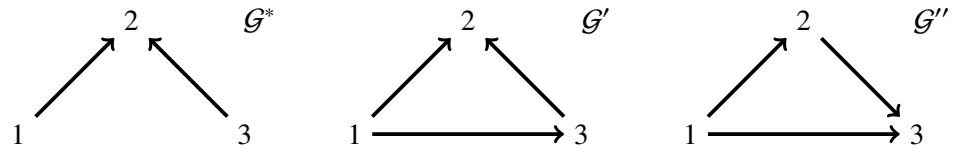


图 8： $\mathcal{G}^*$ 是忠实于 $\mathcal{P}$ 的唯一有向无环图，其中 $\mathrm{I}(\mathcal{P}) = \{\langle \mathrm{X}_1, \mathrm{X}_3 \rangle\}$。此外， $\mathcal{G}', \mathcal{G}'' \in \mathrm{CMC}(\mathcal{P}) \setminus \mathrm{PM}(\mathcal{P})$ 。

令 $\mathcal{P}$ 为定义在 $\mathrm{V} = \{\mathrm{X}_1, \mathrm{X}_2\}$ 上的联合概率分布。考虑 $\mathcal{G}^* \mathrm{DAG(v)}$ 中的 $\mathcal{G}' \in$ ，其中 $\mathcal{G}^*$ 为空有向无环图，$\mathcal{G}'$ 为单边有向无环图。鉴于 $\mathrm{I}(\mathcal{P}) = \{\langle \mathrm{X}_1, \mathrm{X}_2 \rangle\}$，我们有 $\mathcal{G}' \in \mathrm{OriF}(\mathcal{P}) \cap \mathrm{TriF}(\mathcal{P})$ 为空成立。但 $\mathcal{G}' \notin$ SGS $\mathrm{M}(\mathcal{P})$，因为 $\mathcal{G}'$ 是 $\mathcal{G}^* \in \mathrm{CFC}(\mathcal{P})$ 的超图。□

我们借助下面一个简单的集合论事实，总结表1中十一种结构因果剃刀的 *[*类*]*之间所有可能的两两子集关系。

事实4.10。考虑任意两个集合 A 和 B，以及任意 $\mathrm{A}^+ \supseteq \mathrm{A}$ 和 $\mathrm{B}^- \subseteq \mathrm{B}$。若 $\mathrm{A} \backslash \mathrm{B} \quad \varnothing$，则 $\mathrm{A}^+ \backslash \mathrm{B}^-, \varnothing$。

现在，我们引入一个新概念，将十一种因果剃刀二分。回想一下，如果因果剃刀所对应的 *[*类*]* 为空集，那么就能检测到对该剃刀的违反。请注意，某些因果剃刀永远不会被检测到违反。例如，对于任意联合概率分布 $\mathcal{P}$，$\mathrm{Fr}(\mathcal{P})$必然非空，因为总存在一个最稀疏马尔可夫有向无环图。若因果剃刀所对应的 *[*类*]* 永不为空，我们就称该因果剃刀始终可实现。因此，若存在一个联合概率分布，使相应的 *[*类*]* 为空，则该因果剃刀并非始终可实现。换言之，若某因果剃刀并非始终可实现，那么它就会先验地排除某些联合概率分布，因为没有任何 DAG 假设能相对于该分布满足 *[*性质*]* 。现在，我们可以根据这一性质对这十一种因果剃刀进行划分。

定理 4.11。以下陈述成立：

(a) C MC、原始忠实性、SGS-极小性、P-极小性和简约性始终是可实现的 .

(b) CFC、u-P-最小性、关系忠实性、邻接忠实性、三重忠实性和 u-简约性并不总是可实现的。

证明。对于 (a)，对于任意联合概率分布 $\mathcal{P}$ ，$\mathrm{OriF}(\mathcal{P})$ 必然非空，因为完全有向无环图不包含非屏蔽三元组。(a) 中其他因果剃刀始终可实现这一点显而易见。对于 (b)，例 4.4 是 $\mathrm{TriF}(\mathcal{P}) = \varnothing$ 的情形。例 4.5 展示了 $\mathrm{uFr}(\mathcal{P}) = \varnothing$ 的情形。(b) 中其他因果剃刀并非始终可实现，这一点由定理 4.1 得出。□

现在，我们将表1中从 *[*类*]'*推广到 *[*剃刀*]'*。考虑任意一对因果剃刀。令 $\mathrm{r}_1(\cdot)$ 和 $\mathrm{r}_2(\cdot)$ 分别为它们各自的 *[*类*]*, , 令剃刀$_1$ 和剃刀$_2$ 分别为它们各自的 *[*剃刀*]*。剃刀 *['*的逻辑强度可以表述如下：

- 剃刀1与剃刀2逻辑等价，当且仅当对于任意联合概率分布 $\mathcal{P}$，$\mathrm{r1}(\mathcal{P}) = \mathrm{r2}(\mathcal{P})$。
- 若对于任意联合概率分布 $\mathcal{P}$，$\mathrm{r1}(\mathcal{P}) \subseteq \mathrm{r2}(\mathcal{P})$成立，但对于某个联合概率分布 $\mathcal{P}'$，$\mathrm{r2}(\mathcal{P}') \backslash \mathrm{r1}(\mathcal{P}') \varnothing$ 成立，则剃刀1严格强于剃刀2。
- 若存在两个联合概率分布 $\mathcal{P}$和 $\mathcal{P}'$，使得 $\mathrm{r2}(\mathcal{P}) \backslash \mathrm{r1}(\mathcal{P}) \varnothing$ 且 $\mathrm{r1}(\mathcal{P}') \backslash \mathrm{r2}(\mathcal{P}') \quad \varnothing$，则剃刀1与剃刀2在逻辑上相互独立。

我们在图9中展示了这十一种结构因果剃刀之间的逻辑层级。为简化表述，我们在标注每个 *[剃刀]* 时省略“因果”“条件”或“假设”一词。逻辑等价的*[剃刀]*合并为一个节点。每条箭头表示，尾节点处的 *[剃刀]* 严格强于头节点处的 *[剃刀]* 。任何两个节点之间若不存在由箭头构成的单向路径，则它们在逻辑上相互独立。

我们以指出拓展这一比较分析的可能性来结束本节。首先，正如前文所述，我们可以纳入近期提出的因果剃刀，扩充清单。其次，成对比较并不局限于研究子集关系。Zhalama 等人 [2019] 引入了保守弱化关系，以探讨在较强的因果剃刀成立时，如何合理地假设较弱的因果剃刀。[24] 第三，也可以对因果剃刀进行三者比较。例如，[Zhang 2013] 的研究结果表明，SGS-极小性与三重忠实性的合取蕴含 P-极小性。因此，有兴趣的读者可以进一步拓展上述比较分析，使其更加全面。下一节中，我们将引入一种不同类型的因果剃刀——参数最小性，以拓展目前的讨论。

5 参数最小性

本节主要讨论一对根据因果模型的参数数量定义的因果剃刀。读者很快就会发现，定义这两种因果剃刀的任务与第3节中对因果剃刀所作的模糊分类并不一致。首先，它们并非纯粹的参数型概念，因为我们需要利用给定有向无环图的图结构特征 $\mathcal{G}$ 来计算 $|\text{param}(\mathcal{G})|$。另一方面，它们也并非纯粹的结构型概念，因为要得到 $\text{param}(\mathcal{G})$，还需要明确联合概率分布所采用的参数假设。这种半结构性质或许可以解释，为什么参数最小性在文献中鲜少受到关注。

然而，在统计学和计算机科学的大多数（如果不是全部）建模方法中，参数数量都是衡量模型统计复杂度的重要指标。此外，这一概念与因果搜索算法密切相关。例如，一致性评分准则，如 [Haughton, 1988]并将在下一节中进一步讨论，已被研究人员广泛用于评估某个DAG假设与给定观测数据集的拟合程度。简而言之，一致性评分准则（例如， [Schwarz, 1978]中的贝叶斯信息准则）要求参数较少的马尔可夫DAG获得更高的分数。因此，不难看出，以参数数量定义的因果剃刀确实具有一定的合理依据。下面直接给出参数最小性及其独特变体的形式化定义。

定义 5.1。（参数最小性）对于任意联合概率分布 $\mathcal{P}$，

- *[类]* $\text{ParamM}(\mathcal{P}) := \{\mathcal{G} \in \text{CMC}(\mathcal{P}) : \neg\exists \mathcal{G}' \in \text{CMC}(\mathcal{P})$ 使得 $|\text{param}(\mathcal{G}')| < |\text{param}(\mathcal{G})|\}$;
- *[有向无环图]* 的 $\mathcal{G}$ 性质 $\mathcal{P}$是参数最小的（对于 $\mathcal{G} \in \text{ParamM}(\mathcal{P})$而言）；
- *[剃刀]* $(\mathcal{G}^*, \mathcal{P})$满足因果参数最小性假设，当且仅当 $\mathcal{G}^* \in \text{ParamM}(\mathcal{P})$。

[24]用我们的语言来说，如果对于满足剃刀$_1$ 的任意联合概率分布$_2$, , 剃刀$_1$严格强于剃刀$\mathcal{P}$，且 r = ($_2$) $\mathcal{P}$r $\mathcal{P}$ ($_1$)，那么剃刀$_2$ 就是剃刀$_1$的保守弱化。例如，简约性是 u-简约性的保守弱化。

	CFC/uPM	ResF	AdjF	OriF	uFr	Fr	PM	SGSM	TriF	CMC
CFC/uPM		**4.1**	**4.1**	**4.1**	**4.1**	**4.1**	**4.1**	**4.1**	**4.1**	**3.2**
ResF	**4.3**		**4.1**	**4.1**	**4.1**	**4.1**	**4.1**	**4.1**	**4.1**	**3.5**
AdjF	4.6(b)†	4.6(b)†		4.6(b)	**4.7**	**4.1**	**4.1**	**4.1**	**4.1**	**3.3**
OriF	4.5(e)†	4.5(e)†	4.5(e)†		4.5(c)	4.9(a)†	4.9(a)†	4.9(a)	4.5(e)	**3.4**
uFr	4.4(a)†	4.4(a)†	4.4(a)†	4.6(a)		**3.7**	**4.1**	**4.1**	4.4(a)	**3.7**
Fr	4.5(a)†	4.5(a)†	4.4(a)†	4.6(a)†	4.5(a)		**4.1**	**4.1**	4.4(a)†	**3.6**
PM	4.5(a)†	4.5(a)†	4.4(a)†	4.6(a)†	4.5(a)†	4.6(c)		**4.1**	4.4(a)†	**3.9**
SGSM	4.5(a)†	4.5(a)†	4.4(a)†	4.6(a)†	4.5(a)†	4.6(c)†	4.8(b)		4.4(a)†	**3.8**
TriF	4.5(b)†	4.5(b)†	4.5(d)	4.6(b)†	4.5(b)	4.9(b)†	4.9(b)†	4.9(b)		**3.11**
CMC	4.5(a)†	4.5(a)†	4.4(a)†	4.6(a)†	4.5(a)†	4.6(c)†	4.6(a)†	4.8(a)	4.4(a)†	

表1： 以下十一种结构因果剃刀的 *[类]*之间的两两比较。 每个单元格表示该行中的 *[类]* 是否为该列中的 *[类]* 的子集。对角线上的单元格以黑色填充，表示其显然成立。以 青色 填充的单元格表示子集关系根据 定理4.1 或定义成立。以 洋红色 填充的单元格表示 例 4.3–4.9中的子集关系反例。上标 † 表示应用了 事实4.10 与 定理4.1。

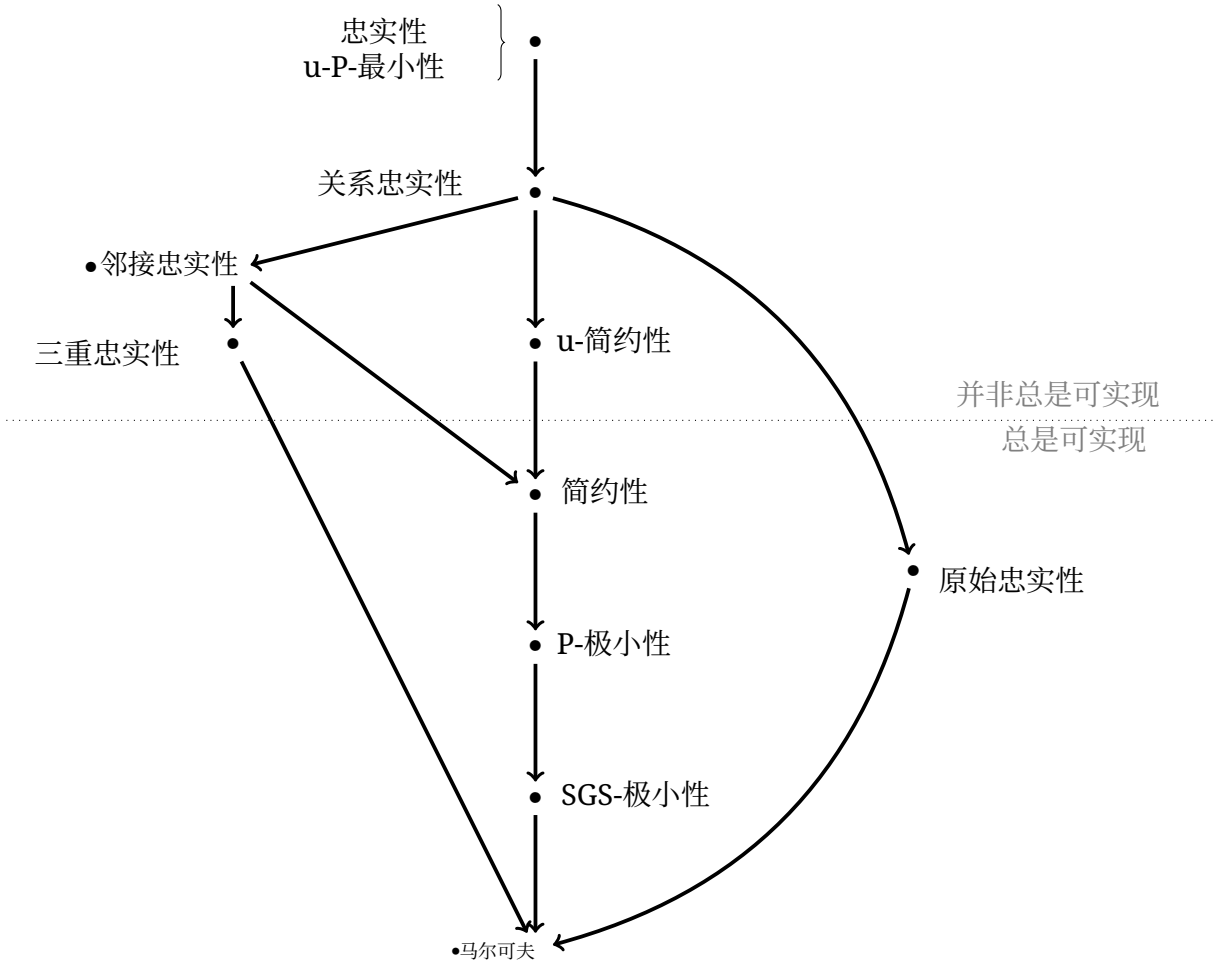


图9：十一种结构因果剃刀的逻辑层级

定义 5.2。（唯一参数极小性）对于任意联合概率分布 $\mathcal{P}$,

- *[类]* $\text{uParamM}(\mathcal{P}) := \begin{cases} \text{ParamM}(\mathcal{P}), & \text{如果 ParamM}(\mathcal{P}) \text{ 中的所有有向无环图都属于同一个马尔可夫等价类} \\ \varnothing ; & \text{否则;} \end{cases}$
- *[性质]* 有向无环图 $\mathcal{G}$ 是 u-参数最小的（相对于 $\mathcal{P}$），当且仅当 $\mathcal{G} \in \text{uParamM}(\mathcal{P})$;
- *[剃刀]* $(\mathcal{G}^*, \mathcal{P})$ 满足因果唯一参数极小性假设，当且仅当 $\mathcal{G}^* \in \text{uParamM}(\mathcal{P})$。

仔细观察后，读者很容易发现，这两种因果剃刀与第3节中定义的简约性和 u-简约性颇为相似。除了上述这对剃刀是根据参数数量而非边数定义之外，它们在各个方面都相似。对于线性高斯因果模型，由于对任意有向无环图 $\mathcal{G}$，等式 $|\text{param}(\mathcal{G})| = |E(\mathcal{G})|$ 成立，因此这两个概念之间的差异不复存在。但在多项式因果模型中，该等式并不成立。

仅关注多项因果模型时，参数最小性及其唯一变体与第3节讨论的其他因果剃刀有何关系？我们首先展示一些关于其 *[类]*的子集关系的正面结论。具体而言，我们将利用 [Chickering、 1995] 和 [Chickering 2002] 关于可覆盖边反转的著名结果，证明对于任意联合多项分布 $\mathcal{P}$，$\text{ParamM}(\mathcal{P}) \subseteq \text{PM}(\mathcal{P})$。

定理 5.3. 对于任意联合多项分布 $\mathcal{P}$，$\text{ParamM}(\mathcal{P}) \subseteq \text{PM}(\mathcal{P})$成立 *s.*

证明。参见附录C。 □

根据定理 4.2 中的 $\text{CFC}(\mathcal{P}) = \text{uPM}(\mathcal{P})$，可以很容易地推出 $\text{CFC}(\mathcal{P}) \subseteq \text{uParamM}(\mathcal{P})$。尽管如此，一个更有意思的问题是，更强的结论 $\text{ResF}(\mathcal{P}) \subseteq \text{uParamM}(\mathcal{P})$ 是否成立。我们利用 [Hemmecke 等人 2012]提出的参数化集合概念来证明这一结论。简言之，我们证明，受限忠实有向无环图必须具有一个最小的参数化集合，并且这意味着该图的参数数量在所有马尔可夫性有向无环图中最少。

定理 5.4. 对于任意联合多项分布 $\mathcal{P}$，$\text{ResF}(\mathcal{P}) \subseteq \text{uParamM}(\mathcal{P})$ 成立。

证明。见附录D。 □

推论 5.5. 对于任意联合多项分布 $\mathcal{P}$，以下陈述为真 *e:*

$$\text{CFC}(\mathcal{P}) \subseteq \text{ResF}(\mathcal{P}) = \text{AdjF}(\mathcal{P}) \cap \text{OriF}(\mathcal{P}) \subseteq \text{uParamM}(\mathcal{P}) \subseteq \text{ParamM}(\mathcal{P}) \subseteq \text{PM}(\mathcal{P}) \subseteq \text{SGSM}(\mathcal{P}) \subseteq \text{CMC}(\mathcal{P}).$$

正面结果到此为止。现在我们转向一些反例，分别针对每个子集关系 $r_1(\cdot) \subseteq r_2(\cdot)$，其中 r_1 和 r_2 中有一个是（u-）参数最小性 *[类]* 。每个反例均来自上一节，该节未涉及参数假设。鉴于当前语境，我们需要说明如何以参数化方式构造底层的多项分布。为简化论述，所有构造细节均留在附录 E。

例 5。 6. 存在一个联合多项分布 $\mathcal{P}$，使得 $\text{AdjF}(\mathcal{P}) \setminus \text{ParamM}(\mathcal{P}), \varnothing.$

[与例 4.7 和图 7 相同。详情请参见附录 E.1。]

首先，$\mathcal{G}_1 \in \text{AdjF}(\mathcal{P})$ 显然成立。鉴于 V 的取值范围为 $\langle 2, 3, 2\rangle$，我们有 $|\text{param}(\mathcal{G}_0)| = 8 > 10 = |\text{param}(\mathcal{G}_1)|$，使得 $\mathcal{G}_1 \notin \text{ParamM}(\mathcal{P})$。□

例 5.7。存在一个联合多项分布 $\mathcal{P}$，使得

(a) $\text{ParamM}(\mathcal{P}) \setminus \text{uParamM}(\mathcal{P}) \neq \emptyset$;

(b) $\text{TriF}(\mathcal{P}) \setminus \text{uParamM}(\mathcal{P}) \neq \emptyset$;

(c) $\text{OriF}(\mathcal{P}) \setminus \text{uParamM}(\mathcal{P}) \neq \emptyset$。

[与例 4.5 和图 5 相同。详情请参见附录 E.2。]

如构造中所指定，V 的取值范围为 $\langle 2、2、2、3\rangle$。我们有 $|\text{param}(\mathcal{G}^*)| = |\text{param}(\mathcal{G}')| = 13$。读者可以验证，$\mathcal{G}^*, \mathcal{G}' \in \text{ParamM}(\mathcal{P})$。因此，$\text{uParamM}(\mathcal{P}) = \emptyset$。如例 4.5 中所述，我们还有 $\mathcal{G}^* \in \text{OriF}(\mathcal{P}) \cap \text{TriF}(\mathcal{P})$。□

推论 5.8。u-参数极小性并非总是可实现的。

例 5.9。存在一个联合多项分布 $\mathcal{P}$，使得

(a) $\text{uParamM}(\mathcal{P}) \setminus \text{OriF}(\mathcal{P}) \neq \emptyset$;

(b) $\text{PM}(\mathcal{P}) \setminus \text{ParamM}(\mathcal{P}) \neq \emptyset$。

[与图 6 中的例 4.6 相同。详情请参见附录 E.3。]

考虑 V 中每个变量都是二元的情况。读者可以验证，$\mathcal{G}^* \in \text{uParamM}(\mathcal{P})$（其中$|\text{param}(\mathcal{G}^*)| = 9$）以及 $\mathcal{G}' \notin \text{ParamM}(\mathcal{P})$（其中 $|\text{param}(\mathcal{G}')| = 13$）。有关$\mathcal{G}^* \notin \text{OriF}(\mathcal{P})$和 $\mathcal{G}' \in \text{PM}(\mathcal{P})$的解释，请参见例 4.6。□

例 5.10. 存在一个联合多项分布 $\mathcal{P}$，使得

(a) $\text{uParamM}(\mathcal{P}) \setminus \text{TriF}(\mathcal{P}) \neq \emptyset$;

(b) $\text{uFr}(\mathcal{P}) \setminus \text{ParamM}(\mathcal{P}) \neq \emptyset$;

(c) $\text{uParamM}(\mathcal{P}) \setminus \text{Fr}(\mathcal{P}) \neq \emptyset$。

[与例 4.4 类似。详情请参见附录 E.4。]

考虑图10中的 $\mathcal{G}_0$ 和 $\mathcal{G}_1$，它们对于 $\mathcal{P}$都满足马尔可夫性。按照构造的设定，V 的取值范围为 $\langle 2、2、2、2、3\rangle$。如例 4.4 所述，我们知道 $\mathcal{G}_0 \in \text{uFr}(\mathcal{P})$。然而，我们有$|\text{param}(\mathcal{G}_0)| = 37 < 35 = |\text{param}(\mathcal{G}_1)|$。事实上，读者可以验证 $\text{uParamM}(\mathcal{P}) = \{\mathcal{G}_1\}$。因此，$\mathcal{G}_0 \in \text{uFr}(\mathcal{P}) \setminus \text{ParamM}(\mathcal{P})$，而 $\mathcal{G}_1 \in \text{uParamM}(\mathcal{P}) \setminus \text{Fr}(\mathcal{P})$。最后，由于 $\mathcal{G}'$ 中的三角形 (3, 5, 4)，$\mathcal{G}_1 \notin \text{TriF}(\mathcal{P})$，并且 $\langle X_3, X_4\rangle \in \Psi_{\mathcal{G}_1,\mathcal{P}}$。□

表2和表11通过在多项式情境中纳入参数最小性和u-参数极小性，扩展了第4节的结果。特别是，（u-）节俭性与（u-）参数最小性之间的分歧揭示了因果搜索算法中的一个两难问题：评估一个有向无环图时，究竟应该计算边还是参数？这一方法论问题将在下一节中进行研究。



图10： $\mathcal{G}_0$ 是 u-节俭的，但不是参数最小的；而 $\mathcal{G}_1$ 是 u-参数最小的，但并不简约。

6 算法含义

上一节所 प्रस्तुत的比较分析展示了如何将众多因果剃刀的逻辑强度系统地排列成一个逻辑层级。然而，某些读者可能会认为这一逻辑项目纯属抽象，对因果搜索算法的开发者或用户不会产生任何实质性影响。本节将说明如何反驳这一观点。

首先，回顾 Lam 等人 [2022] 证明的定理 4.2：CFC 与 u-P-最小性在逻辑上等价。这一逻辑发现意味着，某些因果搜索算法不可避免地会面临 P-最小局部最优问题。更具体地说，当存在可检测到的 CFC 违例（即 $\mathrm{CFC}(\mathcal{P}) = \varnothing$）时，必然至少存在两个彼此不具马尔可夫等价关系的 P-最小有向无环图。假设 $\mathcal{G}_0$ 是一个 P-最小有向无环图，且根据预先选定的标准（例如简约性）尤其理想。在这种情况下，如果因果搜索算法仅以找到某个 P-最小有向无环图（或某个 P-最小有向无环图的马尔可夫等价类）为目标，它可能会返回一个依据预先选定标准（例如， $\mathcal{G}_1$ 比 $\mathcal{G}_0$更稠密）而言并非最优的 P-最小有向无环图 $\mathcal{G}_1$ （或 $\mathrm{MEC}(\mathcal{G}_1)$）。此时，我们称该算法陷入了 P-最小局部最优。

此类算法的一个例子是 Solus 等人最近提出的 TSP 算法 [2021]。简言之，TSP 从一个由顶点的任意初始排列所诱导的有向无环图出发，执行贪心搜索（使用 [Chickering, 2002] 中的 Chickering算法）。他们认为，即使没有 CFC，TSP 也能正确识别出边数最少的有向无环图。遗憾的是，他们的说法并不成立，因为所选的初始排列可能会诱导出一个非最优的 P-极小有向无环图（即并非边数最少的图）。换言之，CFC 是保证 TSP 正确性的必要条件。[25] 这表明，对因果剃刀进行审慎的比较分析，有助于因果搜索算法的开发者发现其研究工作的局限性。

此外，CFC 与 u-P-最小性之间的等价性，为因果发现研究者带来了一个棘手的问题。假设 P-极小性如 Zhang [2013]，所建议的那样，是对真实因果机制作出的一项相当安全的假设。那么，当 CFC 明显不成立时，我们至少会面对两个彼此不具马尔可夫等价性的 P-极小有向无环图。那么，是否存在某种原则性理由，使得某个（或某些）P-极小有向无环图优于其余图?

解决这一问题的一种自然方法，是诉诸比 P-极小性更强的因果剃刀，从而有可能缩小假设空间。可以将这一方法形象地理解为沿图11中的层级向上移动。图中指出了两种直接强于 P-极小性的候选原则：节俭性和参数最小性。尽管如上一节所证明的那样，对于线性高斯模型，这两项原则始终会选取同一组有向无环图，但在多项式模型中，它们可能会得出不同结果。回顾例 5.10（及图10），其中 $\mathcal{G}_0$ 和 $\mathcal{G}_1$ 都是 P-极小的。节俭性因 $\mathcal{G}_0$ 更稀疏而偏好它，而参数最小性则因 $\mathcal{G}_1$ 的参数最少而偏好它。

两种因果剃刀在逻辑上相互独立，这使得实现因果搜索算法时面临两难，尤其是旨在识别得分最高的有向无环图的基于评分的方法。给定一个观测数据集 $\mathcal{D}$，采用预先选定评分准则的基于评分的因果搜索算法，旨在返回一个有向无环图或一类有向无环图，使其相对于 $\mathcal{D}$得分最高。正如读者所料，遵循节俭性原则的评分准则（即偏好更稀疏的马尔可夫模型）可能会作出不同的

[25]参见 [Lam 等人，2022] 有关忠实性对于 TSP 正确性而言为何是必要且充分的更详细分析。

	CFC/uPM	ResF	AdjF	OriF	uFr	Fr	uParamM	ParamM	PM	SGSM	TriF	CMC
CFC/uPM		**4.1**	**4.1**	**4.1**	**4.1**	**4.1**	**5.5**	**5.5**	**4.1**	**4.1**	**4.1**	**3.2**
ResF	**4.3**		**4.1**	**4.1**	**4.1**	**4.1**	**5.5**	**5.5**	**4.1**	**4.1**	**4.1**	**3.5**
AdjF	4.6(b)†	4.6(b)†		4.6(b)	**4.7**	**4.1**	5.6(a)†	5.6(a)	**4.1**	**4.1**	**4.1**	**3.3**
OriF	4.5(e)†	4.5(e)†	4.5(e)†		4.5(c)	4.9(a)†	4.9(a)†	4.9(a)†	4.9(a)†	4.9(a)	4.5(e)	**3.4**
uFr	4.4(a)†	4.4(a)†	4.4(a)†	4.6(a)		**3.7**	5.10(b)†	5.10(b)	**4.1**	**4.1**	4.4(a)	**3.7**
Fr	4.5(a)†	4.5(a)†	4.4(a)†	4.6(a)†	4.5(a)		5.10(b)†	5.10(b)†	**4.1**	**4.1**	4.4(a)†	**3.6**
uParamM	5.10(a)†	5.10(a)†	5.10(a)†	5.9(a)	5.10(c)†	5.10(c)		**5.2**	**5.5**	**5.5**	5.10(a)	**5.2**
ParamM	5.10(a)†	5.10(a)†	5.10(a)†	5.9(a)	5.10(c)†	5.10(c)†	5.7(a)		**5.5**	**5.5**	5.10(a)†	**5.1**
PM	4.5(a)†	4.5(a)†	4.4(a)†	4.6(a)†	4.5(a)†	4.6(c)	5.9(b)†	5.9(b)		**4.1**	**4.4**†	**3.9**
SGSM	4.5(a)†	4.5(a)†	4.4(a)†	4.6(a)†	4.5(a)†	4.6(c)†	5.9(b)†	5.9(b)†	4.8(b)		4.4(a)†	**3.8**
TriF	4.5(b)†	4.5(b)†	4.5(d)	4.6(b)†	4.5(b)	4.9(b)†	5.7(b)	4.9(b)†	4.9(b)†	4.9(b)		**3.11**
CMC	4.5(a)†	4.5(a)†	4.4(a)†	4.6(a)†	4.5(a)†	4.6(c)†	5.9(b)†	5.9(b)†	4.6(a)†	4.8(a)	4.4(a)†	

表2：十三种因果剃刀所得 /类/的两两比较。ParamM 和 uParamM 针对多项因果模型定义。着色规则请参见表1的说明。上标 †表示应用事实4.10以及推论 5.5或定理4.1。

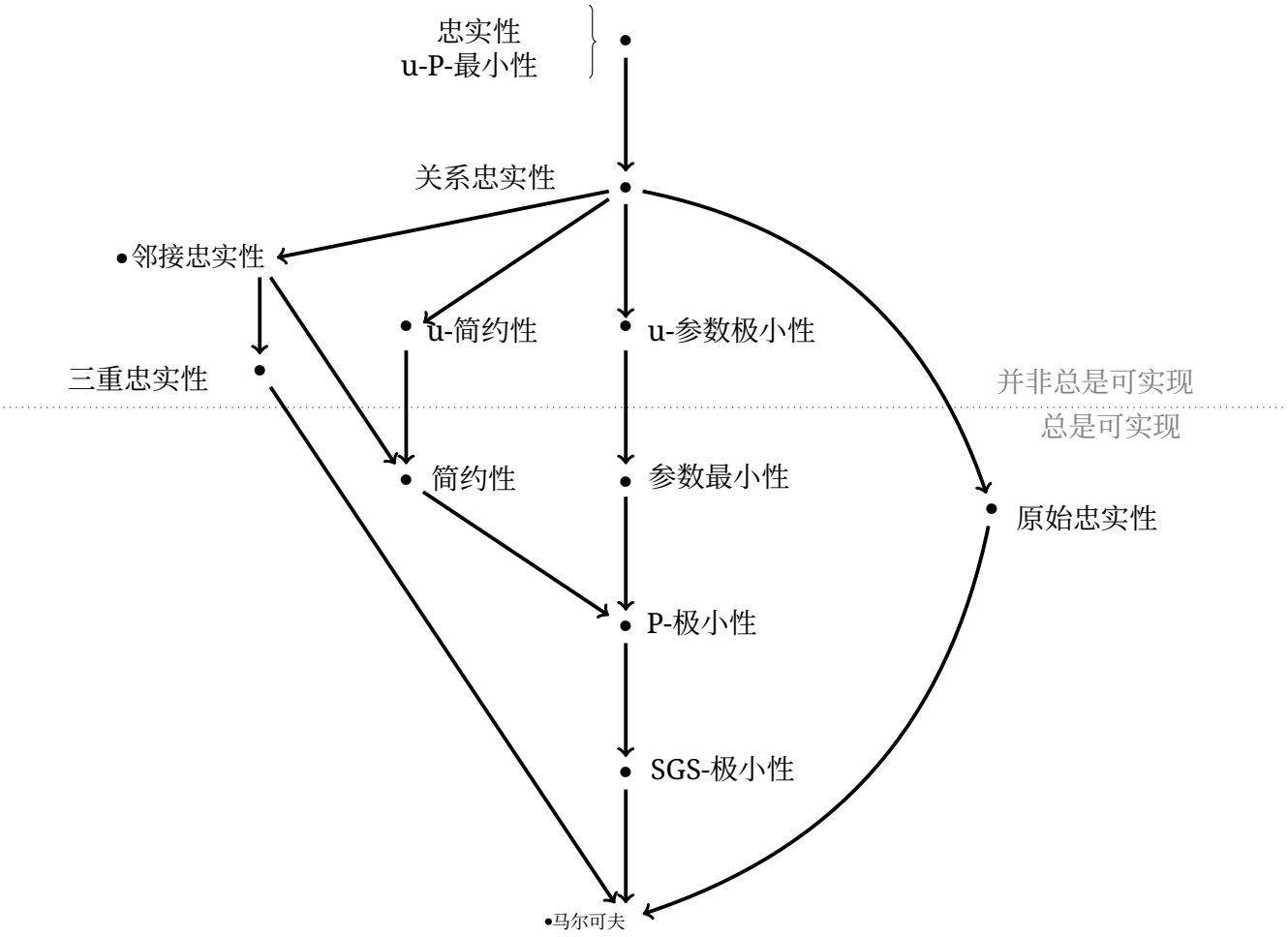


图11：为多项因果模型定义（非）参数最小性时，因果剃刀的逻辑层级。

相较于遵循参数最小性的判断（即偏好参数更少的马尔可夫模型）。因此，两个因果剃刀之间的理论选择如今已转化为一个算法决策。这进一步证实，对因果剃刀的逻辑分析并非纯粹抽象，还会引发实际的两难困境。

让我们进一步深入讨论评分准则。首先，我们假设，每个相关的观测数据集 $\mathcal{D}$ 今后均由变量集 V 上、遵循联合概率分布 $\mathcal{P}$ 的 n 个独立同分布（i.i.d.）样本组成。我们有时写作 $\mathcal{D}_{\mathcal{P}}$，以表示 $\mathcal{D}$ 是从 $\mathcal{P}$ 中抽样得到的。鉴于本节主要在大样本理论的背景下进行研究，暂且假设 n 趋于无穷。

与简约性的本质相一致的一种评分准则，是用于确立 Raskutti 和 Uhler [2018] 以及 Solus 等人 [2021] 所提出的 SP 算法不同变体的负边数。形式上，给定一个关于 V 的观测数据集 $\mathcal{D}_{\mathcal{P}}$，有向无环图 $\mathcal{G} \in \mathrm{DAG(v)}$ 的负边数为：[26]

$$\mathtt{NEC}(\mathcal{G}, \mathcal{D}_{\mathcal{P}}) := \begin{cases} -|\mathtt{E}(\mathcal{G})| & \text{if } \mathcal{G} \in \mathbb{CMC}(\mathcal{P}) \\ -\infty & \textit{otherwise.} \end{cases}$$

显然，NEC 符合简约性原则，即最稀疏的有向无环图必须具有最高评分。

另一方面，文献中最常用的评分准则之一是贝叶斯信息准则（BIC），由 Schwarz 提出 [1978], ，其旨在近似有向无环图的边际对数似然。

定义 6.1.（BIC 评分）给定一个观测数据集 $\mathcal{D}$，其中包含来自联合概率分布 $\mathcal{P}$ 的 n 个独立同分布观测，该分布定义在 V 上，属于曲指数族 [27], ，其中 $\mathrm{X} \in \mathrm{V}$ 且 $\mathrm{S} \subseteq \mathrm{V} \setminus \{\mathrm{X}\}$,

$$\mathtt{BIC}(X, \mathbf{S}, \mathcal{D}_{\mathcal{P}}) = 2\,\ell(\hat{\boldsymbol{\theta}}_{X|\mathbf{S}}\,;\,\mathcal{D}_{\mathcal{P}}) - c\,|\hat{\boldsymbol{\theta}}_{X|\mathbf{S}}|\log(n) \tag{7}$$

ˆ 其中 l 为对数似然函数，$\boldsymbol{\theta}_{X|\mathbf{S}}$ 是以 S 为条件时 X 的最大似然估计量（MLE），而 $\mathrm{c} > 0$是参数惩罚项的乘数（除非另有说明，否则设为 1）。此外，BIC 是一种可分解评分函数，因此 $\mathcal{G} \in$ 有向无环图(v) 的评分（记为 $\mathrm{BIC}(\mathcal{G}, \mathcal{D}_{\mathcal{P}})$）满足以下条件：

$$\begin{aligned} \mathtt{BIC}(\mathcal{G}, \mathcal{D}_{\mathcal{P}}) &= \sum_{i \in \mathbf{v}} \mathtt{BIC}(X_i, \mathbf{X}_{\mathrm{pa}(i,\mathcal{G})}, \mathcal{D}_{\mathcal{P}}) \\ &= \sum_{i \in \mathbf{v}} 2\,\ell(\hat{\boldsymbol{\theta}}_{X_i|\mathbf{X}_{\mathrm{pa}(i,\mathcal{G})}}; \mathcal{D}_{\mathcal{P}}) - c\,|\hat{\boldsymbol{\theta}}_{X_i|\mathbf{X}_{\mathrm{pa}(i,\mathcal{G})}}|\log(n) \\ &= 2\,\ell(\hat{\boldsymbol{\theta}}_{\mathcal{G}}; \mathcal{D}_{\mathcal{P}}) - c\,|\hat{\boldsymbol{\theta}}_{\mathcal{G}}|\log(n) \end{aligned} \tag{8}$$

ˆ ˆ 其中，$\boldsymbol{\theta}_{\mathcal{G}}$ 是对 $\mathcal{G}$ 中独立参数的估计集合，而 $\mathrm{l}(\boldsymbol{\theta}_{\mathcal{G}}; \mathcal{D}_{\mathcal{P}})$ 是 $\mathcal{G}$的最大对数似然估计量（MLLE）。

根据这一表述，对于任意顶点 $\mathrm{i} \in \mathrm{v}$ 和任意 $\mathcal{G} \in$ 有向无环图(v)，在计算 $\mathrm{BIC}(\mathrm{X}_i, \mathrm{X}^{\mathrm{pa}(i,\mathcal{G})}, \mathcal{D}_{\mathcal{P}})$ 时，ˆ MLE $\boldsymbol{\theta}_{X_i|\mathbf{X}^{\mathrm{pa}(i,\mathcal{G})}}$ 是对$\boldsymbol{\theta}_{i,\mathcal{G}}$的估计，也就是说，它是计算 $\mathcal{P}(\mathrm{X}_i \,|\, \mathrm{X}^{\mathrm{pa}(i,\mathcal{G})})$所需的参数集合。ˆ相应地，$\boldsymbol{\theta}_{\mathcal{G}}$ 是在假设后者中的参数彼此独立的情况下，对 $\boldsymbol{\theta}_{\mathcal{G}}$ 的估计。

正如 Haughton [1988],所证明的，贝叶斯信息准则的一个优点是，对于属于曲指数族的分布（包括线性高斯分布和多项分布），它具有一致性。[28]

[26]Solus 等人 [2021] 对负边数的原始定义涉及的是独立性模型 $\mathrm{I}(\mathcal{P})$，而不是具有分布 $\mathcal{P}$的数据集。然而，由于我们假设 $\mathcal{D}_{\mathcal{P}}$ 渐近地服从分布 $\mathcal{P}$，这种表述上的差异不会造成一般性损失。[27]有关曲指数族的深入分析，请参见 [Kass 和 Vos, 2011] 。有关统计模型中曲指数族的更详细探讨，请参见[28]Kass 和 Vos [2011] 。

定义 6.2。（一致性评分准则）给定一个观测数据集 $\mathcal{D}$，其中包含来自定义在 V 上的联合概率分布 $\mathcal{P}$ 的 n 个独立同分布观测值，若在 n 趋于无穷大的极限下，对任意 $\mathcal{G}$、$\mathcal{H} \in$ 有向无环图(v)，评分准则 score 满足以下两个性质，则称其为一致性评分准则：

(a) 如果 $\mathcal{G} \in \text{CMC}(\mathcal{P})$ 且 $\mathcal{H} \notin \text{CMC}(\mathcal{P})$，那么 $\text{score}(\mathcal{G}, \mathcal{D}) > \text{score}(\mathcal{H}, \mathcal{D})$；

(b) 如果 $\mathcal{G}, \mathcal{H} \in \text{CMC}(\mathcal{P})$，且 $|\text{param}(\mathcal{G})| < |\text{param}(\mathcal{H})|$，那么 $\text{score}(\mathcal{G}, \mathcal{D}) > \text{score}(\mathcal{H}, \mathcal{D})$。

定理 6.3. *[*Haughton，1988；Chickering， *2002]* 贝叶斯信息准则是适用于属于曲指数族的联合概率分布的一致性评分准则。

定义 6.2 中的第二个条件与参数最小性的概念十分契合。当一个有向无环图$\mathcal{G}$ 是参数最小的时，所有一致性评分准则（包括贝叶斯信息准则）都会一致判定， $\mathcal{G}$ 在所有有向无环图中得分最高。Chickering [2002] 提出的著名贪婪等价搜索（GES）算法，是一种使用贝叶斯信息准则的经典基于评分的算法。

很容易验证，NEC 仅满足定义 6.2 的第一个条件，而不满足第二个条件。再看一遍例 5.10。我们有 $\text{NEC}(\mathcal{G}_0, \mathcal{D}_{\mathcal{P}}) = -5 > -6 = \text{NEC}(\mathcal{G}_1, \mathcal{D}_{\mathcal{P}})$，而 $|\text{param}(\mathcal{G}_0)| = 37 > 35 = |\text{param}(\mathcal{G}_1)|$（其中 $\mathcal{P}$ 是按照附录 E.4 中的构造得到的多项分布）。这表明，负边数并不是（一致的评分准则）多项分布的一个一致性评分准则。

现在设想一下，你将要在一个观测数据集 $\mathcal{D}_{\mathcal{P}}$ 上部署基于评分的算法，其中 ParamM($\mathcal{P}$)、Fr($\mathcal{P}$)，如例 5.10 所示。你应该选择 NEC 还是 BIC 作为该算法使用的评分准则？你对这个问题的回答反映了你对某种特定方法论因果剃刀的偏好。例如，由于 BIC 得到了广泛应用，看到统计学家和计算机科学家倾向于采用参数最小性也就不足为奇了。一般而言，与非参数最小性假设相比，参数最小假设能够更好地避免有限样本中的过拟合问题。[29] 沿着这一思路，一致性评分准则优于不一致的替代方案。此外，在某些特殊情况下，例如线性高斯模型中，边数或许可以被解释为参数计数的简写。但这种简写并不具有普适性，例如不适用于多项式模型。

不过，上述论证仅仅表明，当目标是解决某个特定问题（例如过拟合）时，应当优先选择一致性评分准则。它并不意味着 NEC 在方法论上本质地劣于一致性评分准则。事实上，不难提出一种偏向 NEC 而非 BIC 的不同目标。例如，计算边数显然比计算 BIC 高效得多。当计算资源不足时，研究人员完全有理由将简约性奉为一种方法论原则，并使用 NEC。

此外，对于底层联合分布无法用参数化方式表示的非参数情形，该怎么办？在这些情形下，分布可能不属于指数曲线族，因此我们没有类似定理 6.3 的一致性结果。不过，可以采用非参数条件独立性检验（例如 [Strobl 等人提出的 KCIT， 2019]）来构建有向无环图，然后使用 NEC 对有向无环图进行评分。因此，NEC 可以视为一种更通用的评分准则，不受任何参数假设的限制。

顺便一提，我们可以从描述性视角探讨这一困境，考察人们在心理层面如何感知因果假设的简洁性。希望这能为哪项原则更具自然主义色彩提供一些线索。心理学领域的多项实验研究，包括 [Bonawitz 和 Lombrozo、 2011] 以及 [Lombrozo， 2012], 都发现了积极证据，表明人们会将因果解释中所援引原因的数量视为简洁性的体现。[30] 用更形式化的方式表述，计算原因的数量可以简化为计算有向无环图中的边数。与此相反，Blanchard 等

[29]正如 MacKay [1992], 所论证的，复杂假设（即参数更多的假设）的边际似然往往更低，因此与简单假设相比，更容易对训练数据过拟合。[30]此外，Pacer 和 Lombrozo [2017] 比较了节点简单性（即解释中涉及的原因数量）与根简单性（即解释中未被解释的原因数量），并观察到解释偏好与后者相一致。

[2018] 提供的证据表明，人们对因果解释简洁性的判断，与在假设中惩罚自由参数的做法一致。显然，目前描述性研究的结果有限，尚不足以解决这一方法论困境。

最后，从认识论视角来看，我们可以质疑简约性或参数最小性哪一种更可能趋近真理。Forster 等人 [2020] 通过论证简约有向无环图蕴含最多的基本条件独立关系，为简约性提供了论据。[31] 然而，他们的论证隐含地假设所有基本条件独立关系的权重相同。在多项式情境下，这一假设值得商榷，因为每个条件独立关系都可以分解为概率等式，而且不同的条件独立关系可能对应数量不同的概率等式。[32] Lam [2023] 表明，参数最小有向无环图必定对应最多的基本概率等式（即基本条件独立关系的概率等式）。另一方面，作者也为参数最小性原则提供了贝叶斯论证。

本文并不追求绝对的判断，而是对评分准则的偏好采取较为灵活的方法论态度。相关的兴趣与目标可能会极大地影响人们在具体情形下对适当评分准则的选择。但最重要的是，算法开发者和使用者应当意识到，他们的选择与对因果剃刀的某种方法论偏好有着微妙的联系。未来的研究将对简约性与参数最小性作更具哲学意味的比较。

7 讨论

因果发现若不对真实因果机制与底层联合概率分布之间的关系作出假设，就无法进行。在本文中，我们以因果剃刀这一通用术语来研究这些假设。文献中讨论的许多因果剃刀都具有结构性，即它们的定义独立于任何算法或参数假设，包括因果忠实性条件和因果简约假设。我们提出了另一类并非纯粹结构性的因果剃刀，即因果参数最小性条件。我们在多项因果模型的背景下，对十三种因果剃刀进行了全面的逻辑比较，并观察到，尽管简约性与参数最小性具有相似之处，但二者在逻辑上相互独立。这引发了算法层面的疑问：在进行因果发现时，无论从方法论还是本体论角度，都应当采用哪一种因果剃刀作为假设。

下面我们简要列出若干拓展因果剃刀研究的方向。首先，我们可以扩展结构因果剃刀的列表。近期的研究，如 [Ng 等人、 2021] 以及 [Zhalama 等人 2019]，提出了一些新的候选者。此外，鉴于附录 D 中对参数化集合的使用，我们可以定义一种结构因果剃刀，将其限制在参数化集合的集合论意义下的极小类中。受引理 D.5 中证明的启发，我们推测，对于多项因果模型而言，这种新的因果剃刀在逻辑上位于受限忠实性与参数最小性之间。

其次，本文没有讨论任何算法因果剃刀。Solus 等人 [2021] 提出的 ESP 假设和 Lam 等人 [2022] 提出的 GRaSP 剃刀分别探讨了这些算法因果剃刀为何严格弱于忠实性假设。它们与本文综述的因果剃刀有何关联，值得进一步研究。此外，我们只考察了线性高斯和多项因果模型背景下的参数最小性（及其唯一变体）。在其他参数假设下（例如条件高斯分布），参数极小性与简约性之间的关系尚不明确。

[31]有向无环图所蕴含的基本条件独立关系集合可以通过有序马尔可夫条件轻松获得。简而言之，给定 π $\mathcal{G}$的拓扑序 $\mathcal{G}$， {⟨ 的基本条件独立关系集合为$_iX_j \mid X_{\mathrm{pa}(i,\mathcal{G})}\rangle$: i ∈ v，且 j 在 π 中先于 i， j ∉ pa(i, $\mathcal{G}$)}。有关该条件的更正式定义，请参见 [Lauritzen 2019]。[32]例如，假设 $X_1, X_2,$ 和 X_3 均为二元变量。那么， $\langle X_1, X_2 \mid \{X_3\}\rangle \in \mathrm{I}(\mathcal{P})$ 可以进一步分解为两个概率等式： $\mathcal{P}(X_1 = 0 \mid X_2 = 0, X_3 = 0) = \mathcal{P}(X_1 = 0 \mid X_2 = 1, X_3 = 0)$，以及 $\mathcal{P}(X_1 = 0 \mid X_2 = 0, X_3 = 1) = \mathcal{P}(X_1 = 0 \mid X_2 = 1, X_3 = 1)$（X X, X）。相反，与边际独立性$\langle X_1, X_2\rangle \in \mathrm{I}(\mathcal{P})$相关的概率等式只有一个： $\mathcal{P}(X_1 = 0 \mid X_2 = 0) = \mathcal{P}(X_1 = 0 \mid X_1 = 1)$。

第三，在实践中，可以根据计算方面的考量来评估应采用哪种因果剃刀。尽管采用较弱的假设总是更稳妥，但弱化忠实性假设也会带来算法上的权衡。以假设忠实性的 GES 为例。尽管忠实性假设经常遭到违反，研究人员仍青睐 GES，因为它具有较高的计算效率和可扩展性。Ramsey 等人 [2017], 的实验表明，某些版本的 GES 甚至能够扩展到一百万个变量。相反，采用 u-简约性时，Raskutti 和 Uhler [2018] 提出的 SP 算法几乎无法处理超过 10 个变量的情况。Lam 等人 [2022] 和 Solus 等人 [2021], 曾探索 SP 的贪心版本，但其可扩展性远不及 GES。这表明，研究人员在决定采用哪种因果剃刀时，实际考量可能会凌驾于理论正确性之上。为了缩小这两方面之间的差距，我们亟需一种快速且可扩展的弱于忠实性假设的算法。

第四，本文对因果剃刀的比较分析也可以通过模拟来研究。例如，我们观察到简约性和参数最小性（及其唯一变体）几乎被违反的频率有多高？如果参数最小性和简约性在样本层面很少出现分歧，那么在有限样本上实现因果搜索算法时，选择假设哪一种因果剃刀，以及选择负边数还是一致性评分准则，都无关紧要。要回答这个问题，就需要开展全面的模拟研究，以验证因果剃刀几乎被违反的频率。

最后，本文讨论的所有因果剃刀都定义在有向无环图上，因此分析局限于因果充分性框架，也就是说，真实因果模型中不存在潜变量。然而，通过在 [Richardson 和 Spirtes中将 d-分离替换为 m-分离， 2002], 我们讨论的所有结构因果剃刀都可以基于能够处理潜变量的极大祖先图（MAG）来定义。采用无环有向混合图（ADMG）的参数化技术，还可以有效地枚举 MAG 中的参数数量（在多项式情境下）。因此，可以合理地预期，我们所有的逻辑结果都可以推广到因果不充分框架。

参考文献

Andersen, H.（2013）。何时应预期因果忠实性遭到违背，以及这为何重要。《科学哲学》，80(5):672–683。Andrews, B., Cooper, G. F., Richardson, T. S., and Spirtes, P.（2022）。无环有向混合图模型的 m-连接整数集合与因子分解。Blanchard, T., Lombrozo, T., and Nichols, S.（2018）。贝叶斯奥卡姆剃刀是大众的剃刀。《认知科学》，42(4):1345–1359。Bonawitz, E. and Lombrozo, T.（2011）。奥卡姆棘轮：儿童如何运用简洁性与概率来约束推断。《发展心理学》，48:1156–64。Chickering, D. M.（1995）。等价贝叶斯网络结构的变换刻画。载于《人工智能不确定性会议论文集》，第87–98页，旧金山，美国，加利福尼亚州。摩根·考夫曼出版社。Chickering, D. M.（2002）。通过贪心搜索进行最优结构识别。《机器学习研究杂志》，3:507–554。Evans, R. J. and Richardson, T. S.（2014）。用于离散数据的马尔可夫无环有向混合图。《统计年鉴》，42(4):1452–1482。

Forster, M., Raskutti, G., Stern, R., and Weinberger, N.（2020）。因果关系的节俭推断。《英国科学哲学杂志》。Haughton, D. M.（1988）。关于如何选择模型来拟合来自指数族的数据。《统计年鉴》，第342–355页。Hemmecke, R., Lindner, S., and Studen´y, M.（2012）。用于学习贝叶斯网络结构的特征 imset。《国际近似推理杂志》，53(9):1336–1349。第五届欧洲概率图模型研讨会（PGM-2010）。Kalisch, M. and B¨uhlmann, P.（2007）。利用 pc-算法估计高维有向无环图。《机器学习研究杂志》，8(22):613–636。Kass, R. E. and Vos, P. W.（2011）。《渐近推断的几何基础》。约翰·威立父子出版公司，霍博肯，新泽西州。Lam, W.-Y.（2023）。《因果剃刀与因果搜索算法》。博士论文，卡内基梅隆大学。Lam, W.-Y., Andrews, B., and Ramsey, J.（2022）。最稀疏排列算法的贪心松弛。载于 Cussens, J. and Zhang, K.（编），《第三十八届人工智能不确定性会议论文集》，《机器学习研究论文集》第180卷，第1052–1062页。PMLR。Lauritzen, S. L.（2019）。图模型讲义。Lin, H. and Zhang, J.（2020）。在不作任何忠实性假设的情况下，从非实验数据中学习因果结构。载于 Kontorovich, A. and Neu, G.（编），《第三十一届国际算法学习理论会议论文集》，《机器学习研究论文集》第117卷，第554–582页。PMLR。Lombrozo, T.（2012）。解释与溯因推理。《牛津思维与推理手册》，第260–276页。MacKay, D. J. C.（1992）。贝叶斯插值。《神经计算》，4(3):415–447。Meek, C.（1995）。利用背景知识进行因果推断与因果解释。载于 UAI。Ng, I., Zheng, Y., Zhang, J., and Zhang, K.（2021）。通过改进精确搜索和较弱假设实现可靠的因果发现。载于 Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W.（编），《神经信息处理系统进展》第34卷，第20308–20320页。Curran Associates, Inc. Pacer, M. and Lombrozo, T.（2017）。奥卡姆剃刀直击根源：因果解释中的简洁性。《实验心理学杂志：综合》，146(12):1761。Pearl, J.（1988）。《智能系统中的概率推理：合理推断网络》。摩根·考夫曼出版社。Pearl, J.（2009）。《因果关系：模型、推理与推断》。剑桥大学出版社，美国，第二版。Ramsey, J., Zhang, J., and Spirtes, P.（2006）。邻接忠实性与保守因果推断。载于《人工智能不确定性会议论文集》，第401–408页。

Ramsey, J. D.、Glymour, M.、Sanchez-Romero, R. 和 Glymour, C.（2017）。百万个变量及更多：用于学习高维图形因果模型的快速贪心等价搜索算法及其在功能性磁共振图像中的应用。《国际数据科学与分析杂志》，3(2):121–129。Raskutti, G. 和 Uhler, C.（2018）。基于最稀疏排列学习有向无环图模型。《统计》，7。Reichenbach, H.（1956）。《时间的方向》。纽约州米尼奥拉：多佛出版社。Richardson, T. 和 Spirtes, P.（2002）。祖先图马尔可夫模型。《统计年鉴》，30(4):962–1030。Schwarz, G.（1978）。估计模型的维度。《统计年鉴》，第461–464页。Shimizu, S.、Hoyer, P. O.、Hyv¨arinen, A.、Kerminen, A. 和 Jordan, M.（2006）。用于因果发现的线性非高斯无环模型。《机器学习研究杂志》，7(10)。Solus, L.、Wang, Y. 和 Uhler, C.（2021）。基于贪心排列的因果推断算法的一致性保证。《生物统计学》，108:795–814。Spirtes, P. 和 Glymour, C.（1991）。一种快速恢复稀疏因果图的算法。《社会科学计算机评论》，9(1):62–72。Spirtes, P.、Glymour, C. 和 Scheines, R.（2000）。《因果、预测与搜索》。麻省理工学院出版社。Spirtes, P. 和 Zhang, J.（2015）。在 k-三角形忠实性假设下因果效应的统一一致估计量。《统计科学》，29。Strobl, E. V.、Zhang, K. 和 Visweswaran, S.（2019）。用于快速非参数因果发现的近似基于核的条件独立性检验。《因果推断杂志》，7(1):20180017。Studen´y, M.（2005）。《概率条件独立结构》。信息科学与统计。施普林格。Uhler, C.、Raskutti, G.、B¨uhlmann, P. 和 Yu, B.（2013）。因果推断中忠实性假设的几何学。《统计年鉴》，第436–463页。Verma, T. 和 Pearl, J.（1988）。因果网络：语义与表达能力。载于《人工智能不确定性会议论文集》，第69–78页。Verma, T. 和 Pearl, J.（1990）。因果模型的等价性与综合。《概率与因果推断》。Weinberger, N.（2017）。忠实性、协调与因果巧合。《认识》。Woodward, J.（2021）。《具有人类面孔的因果关系：规范理论与描述心理学》。牛津大学出版社。Zhalama、Zhang, J.、Eberhardt, F.、Mayer, W. 和 Li, M. J.（2019）。在较弱假设下基于ASP的发现半马尔可夫因果模型。Zhang, J.（2013）。马尔可夫因果模型的三种奥卡姆剃刀比较。《英国科学哲学杂志》，64:423–448。

Zhang, J. 和 Spirtes, P.（2008）。不忠实性检测与稳健因果推断。《思维与机器》，18(2):239–271。

Zhang, J. 和 Spirtes, P.（2011年）。干预、决定论与因果最小性条件。Synthese，182:335–347。

拟阵独立公理

考虑任意两两不相交的变量集合 W、X、Y 和 Z。

$$\begin{aligned}
\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} &\Rightarrow \mathbf{Y} \perp\!\!\!\perp \mathbf{X} \mid \mathbf{Z} && (\textit{symmetry}) \\
\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} &\Rightarrow \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} && (\textit{decomposition}) \\
\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} &\Rightarrow \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \cup \mathbf{W} && (\textit{weak union}) \\
(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}) \wedge (\mathbf{X} \perp\!\!\!\perp \mathbf{W} \mid \mathbf{Z} \cup \mathbf{Y}) &\Rightarrow \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} && (\textit{contraction}) \\
(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \cup \mathbf{W}) \wedge (\mathbf{X} \perp\!\!\!\perp \mathbf{W} \mid \mathbf{Z} \cup \mathbf{Y}) &\Rightarrow \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} && (\textit{intersection}) \\
(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}) \wedge (\mathbf{X} \perp\!\!\!\perp \mathbf{W} \mid \mathbf{Z}) &\Rightarrow \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} && (\textit{composition})
\end{aligned}$$

若分布 $\mathcal{P}$ 的 I($\mathcal{P}$) 在对称性、分解、弱并和收缩下封闭，则称其为半拟阵。若半拟阵 $\mathcal{P}$ 的 I($\mathcal{P}$) 在交集下封闭，则称其为拟阵。若拟阵 $\mathcal{P}$ 的 I($\mathcal{P}$) 在组合下封闭，则称其为组合的。每个联合概率分布 $\mathcal{P}$ 都是半拟阵。每个严格正分布都是拟阵，而每个线性高斯分布都是组合的拟阵。有关拟阵公理的更全面研究，请参见 [Studen ́y，2005 年，第 2] 章。此外，在接下来的证明中，对称性的应用将隐式进行。

B 定理 4.2 的证明

引理 B.1。Lam 等人 *[2022]* 给定 V 上的联合概率分布 $\mathcal{P}$，对于任意 $\langle X_i, X_j \mid X_{\mathbf{s}} \rangle \in I(\mathcal{P})$，存在 $\mathcal{G} \in \mathrm{DAG}(\mathbf{v})$，使得 $I(\mathcal{G}) = \{\langle X_i, X_j \mid X_{\mathbf{s}} \rangle\}$。

证明。假设 V 由 m 个变量组成，其中 $\mathbf{v} = \{1, ..., m\}$，且 $m \geq 2$。当 $m = 2$时，空有向无环图足以证明该引理。现在考虑 $m \geq 3$的情况。假设对于某些互不相同的 $i、j \in \mathbf{v}$ 和 $\mathbf{s} \subseteq \mathbf{v} \setminus \{i, j\}$，有 $\langle X_i, X_j \mid X_{\mathbf{s}} \rangle \in I(\mathcal{P})$。不失一般性，我们重新标记索引，使得 $i = 1$，$j = k + 2$，且 $\mathbf{s} = \langle 2, ..., k + 1 \rangle$，其中 $\mathbf{v} \setminus \mathbf{s} = \langle k + 3, ..., m \rangle$。现在考虑以下算法，其目标是构造所需的 $\mathcal{G} \in \mathrm{DAG}(\mathbf{v})$。

---

1 $\mathcal{G} \leftarrow$ 在 v 上构造一个完全无向图
2 移除邻接关系 $(1, k + 2) \in \mathrm{sk}(\mathcal{G})$
3 对于每个 $(j, k) \in \mathrm{sk}(\mathcal{G})$ 执行
4 如果 $j < k$，则
5 将 $j \mapsto k$ 定向为 $\mathcal{G}$
6 返回 $\mathcal{G}$

第 3 行至第 5 行保证 $\mathcal{G}$是一个有向无环图。这是因为所有边均为有向边，并且都从较低索引指向较高索引，因此不会产生有向环。最后，$1 \perp_{\mathcal{G}} k + 2 \mid \mathbf{s}$ 成立，这是因为从 1 到 k 的所有有向路径 + 2 要么包含一个非碰撞点 $i \in \mathbf{s}$，要么包含一个碰撞点 $i \notin \mathbf{s}$。因此，$I(\mathcal{G}) = \{\langle X_1, X_{k+2} \mid X_{\mathbf{S}} \rangle\}$，因为在 $\mathcal{G}$ 中不存在其他 d-分离关系。□

定理 4.2。Lam 等人 *[2022]* 对于任意联合概率分布 $\mathcal{P}$，$\mathrm{CFC}(\mathcal{P}) = \mathrm{uPM}(\mathcal{P})$。证明。 [⊆]假设 $\mathcal{G} \in \mathrm{CFC}(\mathcal{P})$。由定理4.1(a)可知，$\mathcal{G} \in \mathrm{PM}(\mathcal{P})$。对于任意 $\mathcal{G}' \in \mathrm{CMC}(\mathcal{P})$，如果 $I(\mathcal{G}') \subset I(\mathcal{G})$，那么 $\mathcal{G}' \notin \mathrm{PM}(\mathcal{P})$。因此，如果 $\mathcal{G}' \in \mathrm{PM}(\mathcal{P})$，那么 $I(\mathcal{G}') = I(\mathcal{G})$。因此，$\mathcal{G} \in \mathrm{uPM}(\mathcal{P})$。

[⊇] 假设 $\mathcal{G} \notin \mathrm{CFC}(\mathcal{P})$。根据定义 3.9，$\mathrm{uPM}(\mathcal{P}) \subseteq \mathrm{PM}(\mathcal{P})$，因此只要 $\mathcal{G} \notin \mathrm{PM}(\mathcal{P})$，我们便可立即得到 $\mathcal{G} \notin \mathrm{uPM}(\mathcal{P})$。考虑 $\mathcal{G} \in \mathrm{PM}(\mathcal{P})$ 的情形。由 $\mathcal{G} \notin \mathrm{CFC}(\mathcal{P})$ 可知$\Psi_{\mathcal{G},\mathcal{P}}$, ∅。任取一个不忠实条件独立关系 $\psi \in \Psi_{\mathcal{G},\mathcal{P}}$。根据引理 B.1，我们可以构造一个有向无环图 $\mathcal{G}'$，使得 $I(\mathcal{G}') = \{\psi\}$。因此，存在 $\mathcal{G}'' \in \mathrm{PM}(\mathcal{P})$，使得 $I(\mathcal{G}') \subseteq I(\mathcal{G}'') \subseteq I(\mathcal{P})$。由 $\psi \in I(\mathcal{G}'')$ 但$\psi \notin I(\mathcal{G})$，可知 $\mathcal{G}'' \notin \mathrm{MEC}(\mathcal{G})$。因此，根据 $\mathcal{G}, \mathcal{G}'' \in \mathrm{PM}(\mathcal{P})$，我们得到 $\mathcal{G} \notin \mathrm{uPM}(\mathcal{P})$。□

C 定理 5.3 的证明

我们将利用 [Chickering、 1995] 和 [Chickering 关于可覆盖边反转的著名结果，来证明对于任意联合多项分布 $\mathcal{P}$，ParamM($\mathcal{P}$) $\subseteq$ PM($\mathcal{P}$)。具体而言，对于任意有向无环图 $\mathcal{G}$，若有向边 j $\rightarrow$ k $\in$ E($\mathcal{G}$) 满足 pa(j, $\mathcal{G}$) = pa(k, $\mathcal{G}$) \ {j}，则称其为可覆盖边。

定理 C.1. Chickering *[1995]* 给定变量集合 V，考虑任意一对有向无环图 $\mathcal{G}, \mathcal{H} \in$ DAG(v)，其中 $\mathcal{H} \in$ MEC($\mathcal{G}$)，且 $\mathcal{G}$ 中有 k 条边在 $\mathcal{H}$中的方向相反。那么存在一个由 k 次不同的可覆盖边反转组成的序列 $\mathcal{G}$，使得 $\mathcal{G}$ 在所有反转完成后变为 $\mathcal{H}$。

定理 C.2. Chickering *[2002]* 给定变量集合 V，对于每一对有向无环图 $\mathcal{G}, \mathcal{H} \in$ DAG(v)，如果 I($\mathcal{H}$) $\subseteq$ I($\mathcal{G}$)，则存在一个有向无环图序列，称为 Chickering 序列 $\langle \mathcal{H} = \mathcal{G}_1, \mathcal{G}_2, ..., \mathcal{G}_k = \mathcal{G} \rangle$（从 $\mathcal{H}$ 到 $\mathcal{G}$），使得 I($\mathcal{G}_i$) $\subseteq$ I($\mathcal{G}_{i+1}$)，并且对于每个 $1 \leq$ i $<$ k， $\mathcal{G}_{i+1}$ 均可通过反转一条可覆盖边或删除一条有向边从 $\mathcal{G}_i$ 得到。

引理 C.3. 给定定义在 V 上的联合多项分布 $\mathcal{P}$，任取一对有向无环图 $\mathcal{G}, \mathcal{G}' \in$ DAG(v)，其中 $\mathcal{G}' \in$ MEC($\mathcal{G}$)。则 param($\mathcal{G}$) = param($\mathcal{G}'$)。

证明。根据定理 C.1，只需证明反转一条受保护边后参数数量保持不变。假设 $\mathcal{G}$ 和 $\mathcal{G}'$ 在 DAG(v) 中仅相差一条受保护边的反转，使得对于某些 i, j $\in$ v，E($\mathcal{G}'$) = (E($\mathcal{G}$) \ {i $\rightarrow$ j}) $\cup$ {j $\rightarrow$ i }。进一步记 r(i)、r(j) 和 r(pa(i, $\mathcal{G}$)) 分别为 a、b 和 c。因此，我们有

$$|\theta_{i,\mathcal{G}}| = (a-1)c, \qquad |\theta_{j,\mathcal{G}}| = (b-1)ac, \qquad |\theta_{i,\mathcal{G}'}| = (a-1)bc, \qquad |\theta_{j,\mathcal{G}'}| = (b-1)c.$$

注意，对于每个 l $\in$ v \ {i, j}，$|\theta_{l,\ \mathcal{G}}| = |\theta_{l,\ \mathcal{G}'}|$。由 (4)，我们有：

$$\begin{aligned}
|\text{param}(\mathcal{G})| - |\text{param}(\mathcal{G}')| &= (|\theta_{i,\mathcal{G}}| + |\theta_{j,\mathcal{G}}|) - (|\theta_{i,\mathcal{G}'}| + |\theta_{j,\mathcal{G}'}|) \\
&= \Big((a-1)c + (b-1)ac\Big) - \Big((a-1)bc + (b-1)c\Big) \\
&= c\Big((a-1) + (b-1)a - (a-1)b - (b-1)\Big) \\
&= c\Big((a-1)(1-b) + (b-1)(a-1)\Big) \\
&= 0.
\end{aligned}$$

□

定理 5.3. 对于任意联合多项分布 $\mathcal{P}$，ParamM($\mathcal{P}$) $\subseteq$ PM($\mathcal{P}$) 成立 *s.*

证明。用归谬法，假设 $\mathcal{G} \in$ ParamM($\mathcal{P}$)，但 $\mathcal{G} \notin$ PM($\mathcal{P}$)。因此，存在 $\mathcal{G}' \in$ PM($\mathcal{P}$)，使得 I($\mathcal{G}$) $\subset$ I($\mathcal{G}'$) $\subseteq$ I($\mathcal{P}$)。根据定理 C.2，存在一个 Chickering 序列，可从 $\mathcal{G}$ 得到 $\mathcal{G}'$。根据引理 C.3，可覆盖边反转不会改变参数数量。另一方面，边删除显然会减少参数数量。鉴于 $\mathcal{G}'$ 和 $\mathcal{G}$ 不属于同一个 MEC，必定至少需要执行一次边删除操作，才能从 $\mathcal{G}$得到 $\mathcal{G}'$。换言之，我们有 |param($\mathcal{G}'$) | < |param($\mathcal{G}$)|。这与 $\mathcal{G} \in$ ParamM($\mathcal{P}$)矛盾。□

D 定理 5.10 的证明

首先，我们介绍一些概念，以便证明对于每个联合多项式分布 $\mathcal{P}$，都有 ResF($\mathcal{P}$) $\subseteq$ uParamM($\mathcal{P}$)。具体而言，我们借用了 Evans and Richardson [2014] 中用于参数化无环有向混合图（ADMGs）的参数化集合这一概念。由于有向无环图是无环有向混合图的一个特殊类，这一借用的概念很容易应用于我们的研究主题。[33]

定义 D.1。Studen´y *[2005]* 考虑任意有向无环图 $\mathcal{G} \in$ DAG(v)。特征 imset $c_{\mathcal{G}}$: P(v) \ {$\varnothing$} $\to${0, 1} 定义如下[34]:

$$c_{\mathcal{G}}(\mathbf{s}) := \begin{cases} 1 & \text{if } \{\langle X_i, X_j \mid \mathbf{X_k}\rangle \in \mathfrak{I}(\mathcal{G}) : \mathbf{s} \setminus \mathbf{k} = \{i, j\}\rangle\} = \varnothing \\ 0 & \text{otherwise} \end{cases}$$

对于任意 s $\in$ P(v)，其中 s $\neq \varnothing$。

将 $\mathcal{G}$ 的参数化集合类记为 S($\mathcal{G}$) = {s $\in$ P(v) \ {$\varnothing$} : $c_{\mathcal{G}}$(s) = 1}。换言之，当且仅当在 s 处取值的特征 imset 为 1 时，s 才是参数化集合。下面给出参数化集合的一个很好的图结构特征。

引理 D.2。Hemmecke 等人 *[2012]*；Andrews 等人 *[2022]* 考虑任意有向无环图 $\mathcal{G} \in$ DAG(v) 以及任意 s $\in$ P(v)。则 $c_{\mathcal{G}}$(s) = 1 当且仅当存在 i $\in$ s，使得 s \ {i} $\subseteq$ pa(i, $\mathcal{G}$)。

注意到，上一个引理中的存在性必然是唯一的。也就是说，每个 s $\in$ S($\mathcal{G}$) 都包含一个唯一的顶点 i，使得 s \ {i} $\subseteq$ pa(i, $\mathcal{G}$) 成立。[35] 因此，我们可以按顶点对 S($\mathcal{G}$) 进行划分。定义 S(i, $\mathcal{G}$) 为顶点 i 的参数化集合类，其中 $\in$ v 属于 $\mathcal{G} \in$ DAG(v)，且满足

$$\mathfrak{S}(i, \mathcal{G}) := \{\mathbf{s} \in \mathfrak{S}(\mathcal{G}) : i \in \mathbf{s} \text{ and } \mathbf{s} \setminus \{i\} \subseteq \mathrm{pa}(i, \mathcal{G})\}. \tag{9}$$

因此，$\mathbf{v}$，我们有 S($\mathcal{G}$) = S $_{i \in \mathbf{v}}$ S(i, $\mathcal{G}$)，并且对于任意不同的 i、j $\in$，S(i, $\mathcal{G}$) $\cap$ S(j, $\mathcal{G}$) = $\varnothing$

下面给出一个重要结果，说明如何用另一种方式刻画多项因果模型中有向无环图的参数数量，如式 (4) 所示。

引理 D.3。考虑多项因果模型的任意有向无环图 $\mathcal{G} \in$ DAG(v) 以及任意 i $\in$ v。则

$$\sum_{\mathbf{s} \in \mathfrak{S}(i, \mathcal{G})} \prod_{j \in \mathbf{s}} (\mathrm{r}(j) - 1) = (\mathrm{r}(i) - 1) \times \mathrm{r}(\mathrm{pa}(i, \mathcal{G})). \tag{10}$$

证明。考虑最简单的情形，其中 pa($\mathcal{G}$, i) = $\varnothing$。则 S(i, $\mathcal{G}$) = {{i}}，因此式 (10) 的左侧为 r(i) − 1，这等价于右侧（因为 r($\varnothing$) = 1）。

下文中，不失一般性，我们将 i 记为 0。令 |pa(0, $\mathcal{G}$)| = k $\leq$ |v| − 1，使得 pa(0, $\mathcal{G}$) = {1, ..., k}。为得到 S(0, $\mathcal{G}$)，注意到 S(0, $\mathcal{G}$) 与 P(pa(0, $\mathcal{G}$)) 之间存在双射，因此每个 s $\in$ S(0, $\mathcal{G}$) 都可以由某个 s′ $\cup$ {0} 唯一确定，其中 s′ $\in$ P(pa(0, $\mathcal{G}$))。

[33] 我非常感谢 Bryan Andrews 指出了这一思路以及相关文献，包括 Hemmecke 等人 [2012] 和 Andrews 等人 [2022]。他还勾勒出了上述结论的主要证明。[34] 对于任意集合 X，P(X) 表示 X 的幂集。[35] 用归谬法，假设存在 i、j $\in$ s，使得 s \ {i} $\subseteq$ pa(i, $\mathcal{G}$) 且 s \ { j} $\subseteq$ pa(j, $\mathcal{G}$)。那么 j $\in$ pa(i, $\mathcal{G}$)，且 i $\in$ pa(j, $\mathcal{G}$)。矛盾。

现在，我们要通过对 k 使用归纳法来证明 (10)。首先，考虑 $\mathrm{k} = 1$ 的基础情形，此时 $\mathrm{pa}(0, \mathcal{G}) = \{1\}$ 且 $\mathrm{S}(0, \mathcal{G}) = \{\{0\}, \{0, 1\}\}$。(10) 的左侧变为

$$\sum_{\mathbf{s} \in \mathfrak{S}(0, \mathcal{G})} \prod_{j \in \mathbf{s}} (\mathrm{r}(j) - 1) = (\mathrm{r}(0) - 1) + (\mathrm{r}(0) - 1)(\mathrm{r}(1) - 1) = (\mathrm{r}(0) - 1)\, \mathrm{r}(1)$$

其中 $\mathrm{r}(1) = \mathrm{r}(\{1\}) = \mathrm{r}(\mathrm{pa}(0, \mathcal{G}))$。因此，基础情形得证。考虑归纳假设：对于某个任意的 $\mathrm{k} - 1 \leq |\mathrm{v}| - 2$，若有向无环图 $\mathcal{G}$ 满足 $\mathrm{pa}(0, \mathcal{G}) = \{1, ..., \mathrm{k} - 1\}$，则式 (10) 成立。接着，我们通过添加有向边 $\mathrm{k} \to 0$ to $\mathcal{G}$ 来构造$\mathcal{G}'$，使得 $\mathrm{pa}(0, \mathcal{G}') = \{1, ..., \mathrm{k} - 1, \mathrm{k}\}$。我们要证明，对于 $\mathcal{G}'$ 和顶点 0，式 (10) 成立。

注意，$\mathrm{S}(0, \mathcal{G}) \subset \mathrm{S}(0, \mathcal{G}')$，前者的大小为 2^{k-1}，后者的大小为 2^k。此外，$\mathrm{S}(0, \mathcal{G}') \setminus \mathrm{S}(0, \mathcal{G})$ 与 $\mathrm{S}(0, \mathcal{G})$ 之间存在一个双射，使得每个 $\mathrm{s}' \in \mathrm{S}(0, \mathcal{G}') \setminus \mathrm{S}(0, \mathcal{G})$ 都能唯一地表示为 $\mathrm{s} \cup \{\mathrm{k}\}$，其中某个 $\mathrm{s} \in \mathrm{S}(0, \mathcal{G})$。由此可得如下结论。

$$\sum_{\mathbf{s} \in \mathfrak{S}(0, \mathcal{G}')} \prod_{j \in \mathbf{s}} (\mathrm{r}(j) - 1) = \sum_{\mathbf{s} \in \mathfrak{S}(0, \mathcal{G})} \prod_{j \in \mathbf{s}} (\mathrm{r}(j) - 1) + \sum_{\mathbf{s} \in \mathfrak{S}(0, \mathcal{G}') \setminus \mathfrak{S}(0, \mathcal{G})} \prod_{j \in \mathbf{s}} (\mathrm{r}(j) - 1) \tag{11}$$

特别考虑 (11) 右侧的第二项。

$$\begin{aligned}
\sum_{\mathbf{s} \in \mathfrak{S}(0, \mathcal{G}') \setminus \mathfrak{S}(0, \mathcal{G})} \prod_{j \in \mathbf{s}} (\mathrm{r}(j) - 1) &= \sum_{\mathbf{s} \in \mathfrak{S}(0, \mathcal{G}') \setminus \mathfrak{S}(0, \mathcal{G})} \prod_{j \in \mathbf{s} \setminus \{k\}} (\mathrm{r}(j) - 1)(\mathrm{r}(k) - 1) \\
&= (\mathrm{r}(k) - 1) \times \sum_{\mathbf{s} \in \mathfrak{S}(0, \mathcal{G}') \setminus \mathfrak{S}(0, \mathcal{G})} \prod_{j \in \mathbf{s} \setminus \{k\}} (\mathrm{r}(j) - 1) \\
&= (\mathrm{r}(k) - 1) \times \sum_{\mathbf{s} \in \mathfrak{S}(0, \mathcal{G})} \prod_{j \in \mathbf{s}} (\mathrm{r}(j) - 1) \\
&= (\mathrm{r}(k) - 1) \times (\mathrm{r}(0) - 1) \times \mathrm{r}(\mathrm{pa}(0, \mathcal{G}))
\end{aligned}$$

其中，最后一个等式由归纳假设得出。最后，将上述结果代入 (11)：

$$\begin{aligned}
\sum_{\mathbf{s} \in \mathfrak{S}(0, \mathcal{G}')} \prod_{j \in \mathbf{s}} (\mathrm{r}(j) - 1) &= \sum_{\mathbf{s} \in \mathfrak{S}(0, \mathcal{G})} \prod_{j \in \mathbf{s}} (\mathrm{r}(j) - 1) + \sum_{\mathbf{s} \in \mathfrak{S}(0, \mathcal{G}') \setminus \mathfrak{S}(0, \mathcal{G})} \prod_{j \in \mathbf{s}} (\mathrm{r}(j) - 1) \\
&= \Big((\mathrm{r}(0) - 1) \times \mathrm{r}(\mathrm{pa}(0, \mathcal{G})) \Big) + \Big((\mathrm{r}(k) - 1) \times (\mathrm{r}(0) - 1) \times \mathrm{r}(\mathrm{pa}(0, \mathcal{G})) \Big) \\
&= (\mathrm{r}(0) - 1) \times \mathrm{r}(\mathrm{pa}(0, \mathcal{G})) \times (1 + (\mathrm{r}(k) - 1)) \\
&= (\mathrm{r}(0) - 1) \times \mathrm{r}(\{1, ..., k - 1\}) \times \mathrm{r}(k) \\
&= (\mathrm{r}(0) - 1) \times \mathrm{r}(\{1, ..., k - 1, k\}) \\
&= (\mathrm{r}(0) - 1) \times \mathrm{r}(\mathrm{pa}(0, \mathcal{G}')).
\end{aligned}$$

□

推论 D.4. 考虑多项因果模型的任意有向无环图 $\mathcal{G} \in \mathrm{DAG}(\mathrm{v})$。则

$$|\mathtt{param}(\mathcal{G})| = \sum_{i \in \mathbf{v}} \sum_{\mathbf{s} \in \mathfrak{S}(i, , \mathcal{G})} \prod_{j \in \mathbf{s}} (\mathrm{r}(j) - 1) = \sum_{\mathbf{s} \in \mathfrak{S}(\mathcal{G})} \prod_{j \in \mathbf{s}} (\mathrm{r}(j) - 1). \tag{12}$$

现在我们证明一个关键引理：受限忠实有向无环图中的参数化集合类必然是任意马尔可夫性有向无环图中相应集合类的子集。因此，这一结论推出了我们所需的结果：对于任意联合多项分布 $\mathcal{P}$，$\mathrm{ResF}(\mathcal{P}) \subseteq \mathrm{uParamM}(\mathcal{P})$。

引理 D.5。对于任意联合多项分布 $\mathcal{P}$，若 $\mathcal{G} \in \text{ResF}(\mathcal{P})$，则对于任意 $\mathcal{G}' \in \text{CMC}(\mathcal{P})$，$\text{S}(\mathcal{G}) \subseteq \text{S}(\mathcal{G}')$。

证明。首先，根据 $\mathcal{G} \in \text{AdjF}(\mathcal{P})$，对于任意 $\mathcal{G}' \in \text{CMC}(\mathcal{P})$，我们有 $\text{sk}(\mathcal{G}) \subseteq \text{sk}(\mathcal{G}')$。现在，我们通过考察每个 $\text{s} \in \text{S}(\mathcal{G})$ 的基数来证明该引理（其中 $|\text{s}| \geq 1$，如定义 D.1 所要求）。当 $|\text{s}| = 1$ 时，显然有 $\text{s} \in \text{S}(\mathcal{G}')$。考虑 $|\text{s}| = 2$，其中 $\text{s} = \{\text{i}, \text{j}\}$，且 $\text{i}, \text{j} \in \text{v}$ 互不相同。注意，s 对应于 $\mathcal{G}$ 中 i 与 j 之间的邻接关系。由于 $\text{sk}(\mathcal{G}) \subseteq \text{sk}(\mathcal{G}')$，i 和 j 在 $\mathcal{G}'$ 中也必须相邻，因此 $\{\text{i}, \text{j}\} = \text{s} \in \text{S}(\mathcal{G}')$。

考虑 $|\text{s}| = 3$，其中 $\text{s} = \{\text{i}, \text{j}, \text{k}\}$。鉴于 $\text{s} \in \text{S}(\mathcal{G})$，根据引理 D.2，$\{\text{i}, \text{j}, \text{k}\}$ 在 $\mathcal{G}$ 中要么形成无屏蔽碰撞，要么形成三角形（因为在有向无环图中，为避免出现环，三角形必须是有屏蔽碰撞）。根据 $\mathcal{G} \in \text{ResF}(\mathcal{P})$，$\{\text{i}, \text{j}, \text{k}\}$ 在 $\mathcal{G}'$ 中也必须形成无屏蔽碰撞或三角形。因此，根据引理 D.2，$\{\text{i}, \text{j}, \text{k}\} = \text{s} \in \text{S}(\mathcal{G}')$。

最后，考虑 $|\text{s}| > 3$。根据引理 D.2，必定存在 $\text{i} \in \text{s}$，使得 $\text{s} \setminus \{\text{i}\} \subseteq \text{pa}(\text{i}, \mathcal{G})$。因此，对于任意一对不同的 j、$\text{k} \in \text{s} \setminus \{\text{i}\}$，都有 $\{\text{i}, \text{j}\}, \{\text{i}, \text{k}\}, \{\text{i}, \text{j}, \text{k}\} \in \text{S}(\mathcal{G})$。由上述结果可知，我们还得到$\{\text{i}, \text{j}\}, \{\text{i}, \text{k}\}, \{\text{i}, \text{j}, \text{k}\} \in \text{S}(\mathcal{G}')$。这意味着，对于每个 $\text{l} \subseteq \text{v}$（其中 $\text{s} \setminus \text{l} = \{\text{i}, \text{j}, \text{k}\}$），都有 $\langle \text{X}_i, \text{X}_j \,|\, \text{X}_\text{l} \cup \{\text{X}_k\}\rangle, \langle \text{X}_i, \text{X}_k \,|\, \text{X}_\text{l} \cup \{\text{X}_j\}\rangle, \langle \text{X}_j, \text{X}_k \,|\, \text{X}_\text{l} \cup \{\text{X}_i\}\rangle \notin \text{I}(\mathcal{G}')$。根据定义 D.1，可得 $\text{c}_{\mathcal{G}'}(\text{s}) = 1$，因此 $\text{s} \in \text{S}(\mathcal{G}')$。□

定理 5.4. 对任意联合多项分布 $\mathcal{P}$，$\text{ResF}(\mathcal{P}) \subseteq \text{uParamM}(\mathcal{P})$ 成立 *s.*

证明。考虑 $\mathcal{G} \in \text{ResF}(\mathcal{P}) = \text{AdjF}(\mathcal{P}) \cap \text{OriF}(\mathcal{P})$。根据 $\mathcal{G} \in \text{AdjF}(\mathcal{P})$，若有向无环图 $\mathcal{G}' \in \text{DAG(v)}$ 缺少 $\mathcal{G}$ 的任何邻接关系，则 $\mathcal{G}'$（对于 $\mathcal{P}$）不具马尔可夫性。此观察可推出三个结果。首先，没有任何比 $\mathcal{G}$ 更稀疏的有向无环图具备马尔可夫性。其次，对于任意严格稠密于 $\mathcal{G}$ 的马尔可夫 $\mathcal{G}'$，$\mathcal{G}'$ 中的邻接关系集合必定真包含 $\mathcal{G}$ 中的邻接关系集合（即 $\text{sk}(\mathcal{G}) \subset \text{sk}(\mathcal{G}')$）。最后，与 $\mathcal{G}$ 同样稀疏的马尔可夫DAG必须与 $\mathcal{G}$ 具有相同的骨架。最后这一点结合 $\mathcal{G} \in \text{OriF}(\mathcal{P})$，意味着所有这些同样稀疏的有向无环图都与 $\mathcal{G}$ 属于同一个马尔可夫等价类，因此根据引理 C.3，它们具有相同的参数数量。换言之，要证明 $\mathcal{G} \in \text{uParamM}(\mathcal{P})$，只需证明对于任意满足 $\text{sk}(\mathcal{G}) \subset \text{sk}(\mathcal{G}')$ 的 $\mathcal{G}' \in \text{CMC}(\mathcal{P})$，都有 $|\text{param}(\mathcal{G})| < |\text{param}(\mathcal{G}')|$。

现在考虑任意一个这样的 $\mathcal{G}'$。根据引理 D.5，我们有 $\text{S}(\mathcal{G}) \subseteq \text{S}(\mathcal{G}')$。然而，请注意，必然存在 i、$\text{j} \in \text{v}$，使得 $(\text{i}, \text{j}) \in \text{sk}(\mathcal{G}') \setminus \text{sk}(\mathcal{G})$。由此可得，$\{\text{i}, \text{j}\} \in \text{S}(\mathcal{G}')$，但 $\{\text{i}, \text{j}\} \notin \text{S}(\mathcal{G})$。因此，我们有 $\text{S}(\mathcal{G}) \subset \text{S}(\mathcal{G}')$。所以，根据推论 D.4 中的 (12)，$|\text{param}(\mathcal{G})| < |\text{param}(\mathcal{G}')|$。□

E 构造示例

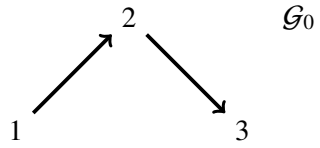
本节讨论如何以构造方式模拟第5节中的例子。回顾一下，这些例子中的$^\circ$主张都是存在一个联合多项分布 $\mathcal{P}$（其中，开销 $\circ$表示我们旨在证明其存在的是联合概率分布），该分布满足某些条件（例如，$\mathrm{AdjF}(\mathring{\mathcal{P}}) \setminus \mathrm{ParamM}(\mathring{\mathcal{P}}) \ \varnothing$）。还要注意，每个例子都涉及一个特定的有向无环图 $\mathcal{G} \in \mathrm{CMC}(\mathring{\mathcal{P}})^\circ$，使得 $\mathrm{I}(\mathcal{P}) \setminus \mathrm{I}(\mathcal{G}) = \Psi_{\mathcal{G},\, \mathring{\mathcal{P}}}, \varnothing$（即，$\mathcal{G}$不忠实于 $\mathcal{P}$）。为了证明这样的 $\mathcal{P}$存在，我们构造一个联合多项分布 $\mathcal{P}$，使得 $\mathrm{I}(\mathcal{P})$满足相关条件。该构造可分为两个步骤。首先，我们通过 d-分离得到 $\mathrm{I}(\mathcal{G})$，并通过对概率进行特定赋值，在 $\Psi_{\mathcal{G},\, \mathring{\mathcal{P}}}$ 中引入条件独立关系集合，使得 $\mathrm{I}(\mathcal{G}) \cup \Psi_{\mathcal{G},\, \mathring{\mathcal{P}}} \subseteq \mathrm{I}(\mathcal{P})$。第二步是确保上述赋值不会在 $\mathrm{I}(\mathcal{P})$ 中引入任何额外的条件独立关系。完成这两个步骤后，我们便得到$^\circ\mathrm{I}(\mathcal{P}) = \mathrm{I}(\mathcal{G}) \cup \Psi_{\mathcal{G},\, \mathring{\mathcal{P}}} = \mathrm{I}(\mathcal{P})$，如所要求。由于第二步涉及繁琐但直接的算术推导，除例 5.6 中作为示例说明外，我们不再展开细节。

现在，我们为构造引入一些必要的记号。考虑定义在 V 上的任意联合多项分布 $\mathcal{P}$。接下来，给定所讨论的有向无环图 $\mathcal{G}$ 与 $\mathcal{P}$满足马尔可夫关系，我们可以用马尔可夫分解表示联合多项分布 $\mathcal{P}$：

$$\mathcal{P}(\mathbf{V}) = \prod_{i \in \mathbf{v}} \mathcal{P}(X_i \mid \mathbf{X}_{\mathrm{pa}(i,\mathcal{G})}). \tag{6}$$

注意，每个 $\mathcal{P}(\mathrm{X}_i \mid \mathrm{X}_{\mathrm{pa}(i,\mathcal{G})})$ 均由参数集合 $\boldsymbol{\theta}_{i,\mathcal{G}}$ 决定。我们将每个 $\boldsymbol{\theta}_{i,\mathcal{G}}$ 以表格形式表示（例如例 5.6 中的三张表），并称之为 $\boldsymbol{\theta}$表。$\boldsymbol{\theta}_{i,\mathcal{G}}$ 的表格维度为 $\mathrm{r}(\mathrm{pa}(\mathrm{i},\mathcal{G})) \times \mathrm{r}(\mathrm{i})$，其中每行对应 $\mathrm{X}_{\mathrm{pa}(i,\mathcal{G})}$的一个取值，每列对应 X_i的一个取值。每个单元格表示在 $\mathrm{X}_{\mathrm{pa}(i,\mathcal{G})}$ 取值为该行指定的值的条件下，X_i 取该列指定值的概率。由于每行的和必须为 1，最后一列中的单元格由同一行其余单元格的值决定。因此，计算 $\mathcal{P}(\mathrm{X}_i \mid \mathrm{X}_{\mathrm{pa}(i,\mathcal{G})})$ 所需的参数不包括最后一列中的单元格，故将其标记为灰色。

E.1 例 5.6



$\mathcal{G}_0$

图12： $\mathcal{G}_0$ 来自图 7

给定 $\mathrm{V} = \{\mathrm{X}_1, \mathrm{X}_2, \mathrm{X}_3\}$，考虑 $\mathrm{r}(1) = \mathrm{r}(3) = 2$ 和 $\mathrm{r}(2) = 3$。我们希望构造一个多项式$^\circ$分布 $\mathcal{P}$，使得 $\mathrm{I}(\mathcal{P}) = \{\langle \mathrm{X}_1, \mathrm{X}_3 \mid \{\mathrm{X}_2\}\rangle, \langle \mathrm{X}_1, \mathrm{X}_3\rangle\} = \mathrm{I}(\mathcal{P})$。首先，我们从图12中的 $\mathcal{G}_0$ 开始，其中 $\mathrm{I}(\mathcal{G}_0) = \{\langle \mathrm{X}_1, \mathrm{X}_3 \mid \{\mathrm{X}_2\}\rangle\}$。以下是这三个变量的 $\boldsymbol{\theta}$-表。

$\boldsymbol{\theta}_{1,\mathcal{G}_0}$

i	0	1
$\mathcal{P}(\mathrm{X}_1 = \mathrm{i})$	$\mathrm{a} = 0.5$	0.5

$\boldsymbol{\theta}_{2,\mathcal{G}_0}$

i	0	1	2
$\mathcal{P}(\mathrm{X}_2 = \mathrm{i} \mid \mathrm{X}_1 = 0)$	$\mathrm{b}_0 = 0.2$	$\mathrm{c}_0 = 0.2$	0.6
$\mathcal{P}(\mathrm{X}_2 = \mathrm{i} \mid \mathrm{X}_1 = 1)$	$\mathrm{b}_1 = 0.3$	$\mathrm{c}_1 = 0.3$	0.4

$\boldsymbol{\theta}_{3,\mathcal{G}_0}$

i	0	1
$\mathcal{P}(\mathrm{X}_3 = \mathrm{i} \mid \mathrm{X}_2 = 0)$	$\mathrm{d}_0 = 0.2$	0.8
$\mathcal{P}(\mathrm{X}_3 = \mathrm{i} \mid \mathrm{X}_2 = 1)$	$\mathrm{d}_1 = 0.4$	0.6
$\mathcal{P}(\mathrm{X}_3 = \mathrm{i} \mid \mathrm{X}_2 = 2)$	$\mathrm{d}_2 = 0.3$	0.7

要验证 $\langle \mathrm{X}_1, \mathrm{X}_3 \rangle \in \mathrm{I}(\mathcal{P})$，只需证明 $\mathcal{P}(\mathrm{X}_3 = 0 \,|\, \mathrm{X}_1 = 0) = \mathcal{P}(\mathrm{X}_3 = 0 \,|\, \mathrm{X}_1 = 1)$。考虑任意 $\mathrm{i} \in \{0, 1\}$。

$$
\begin{aligned}
\mathcal{P}(X_3 = 0 \,|\, X_1 = i) &= \sum_{j=0}^{2} \mathcal{P}(X_3 = 0, X_2 = j \,|\, X_1 = i) = \sum_{j=0}^{2} \frac{\mathcal{P}(X_1 = i, X_2 = j, X_3 = 0)}{\mathcal{P}(X_1 = i)} \\
&= \sum_{j=0}^{2} \frac{\mathcal{P}(X_1 = i)\mathcal{P}(X_2 = j \,|\, X_1 = i)\mathcal{P}(X_3 = 0 \,|\, X_2 = j)}{\mathcal{P}(X_1 = i)} \\
&= \sum_{j=0}^{2} \mathcal{P}(X_3 = 0 \,|\, X_2 = j)\mathcal{P}(X_2 = j \,|\, X_1 = i) \\
&= d_0 b_i + d_1 c_i + d_2(1 - b_i - c_i) \\
\mathcal{P}(X_3 = 0 \,|\, X_1 = 0) &= (0.2 \times 0.2) + (0.4 \times 0.2) + (0.3 \times 0.6) = 0.3 \\
\mathcal{P}(X_3 = 0 \,|\, X_1 = 1) &= (0.2 \times 0.3) + (0.4 \times 0.3) + (0.3 \times 0.4) = 0.3
\end{aligned}
$$

因此 $\langle \mathrm{X}_1, \mathrm{X}_3 \rangle \in \mathrm{I}(\mathcal{P})$。

另一方面，我们通过算术推导表明，$\mathrm{I}(\mathcal{P})$ 中除 $\langle \mathrm{X}_1, \mathrm{X}_3 \rangle$ 和 $\langle \mathrm{X}_1, \mathrm{X}_3 \,|\, \{\mathrm{X}_2\} \rangle$之外不存在其他条件独立关系。忽略对称对后，$\mathrm{I}(\mathrm{V})$ 中总共有 6 个可能的条件独立关系。$\langle \mathrm{X}_1, \mathrm{X}_3 \rangle$、$\langle \mathrm{X}_1, \mathrm{X}_3 \,|\, \{\mathrm{X}_2\} \rangle \in \mathrm{I}(\mathcal{P})$ 已在上文证明。对于 $\langle \mathrm{X}_1, \mathrm{X}_2 \rangle$、$\langle \mathrm{X}_2, \mathrm{X}_3 \rangle \in \mathrm{I}(\mathrm{V})$，读者只需分别查看 $\boldsymbol{\theta}_2$、$\mathcal{G}_0$ 和 $\boldsymbol{\theta}_3$、$\mathcal{G}_0$ 的表，即可验证它们不属于 $\mathrm{I}(\mathcal{P})$。对于 $\langle \mathrm{X}_1, \mathrm{X}_2 \,|\, \{\mathrm{X}_3\} \rangle \in \mathrm{I}(\mathrm{V})$，只需证明 $\mathcal{P}(\mathrm{X}_1 = 0 \,|\, \mathrm{X}_2 = 0, \mathrm{X}_3 = 0) \; \mathcal{P}(\mathrm{X}_1 = 0 \,|\, \mathrm{X}_2 = 1, \mathrm{X}_3 = 0)$。对于任意 $\mathrm{j} \in \{0, 1, 2\}$,

$$
\begin{aligned}
\mathcal{P}(X_1 = 0 \,|\, X_2 = j, X_3 = 0) &= \frac{\mathcal{P}(X_1 = 0, X_2 = j, X_3 = 0)}{\mathcal{P}(X_2 = j, X_3 = 0)} \\
&= \frac{\mathcal{P}(X_1 = 0)\mathcal{P}(X_2 = j \,|\, X_1 = 0)\mathcal{P}(X_3 = 0 \,|\, X_2 = j)}{\mathcal{P}(X_2 = j)\mathcal{P}(X_3 = 0 \,|\, X_2 = j)} \\
&= \frac{\mathcal{P}(X_1 = 0)\mathcal{P}(X_2 = j \,|\, X_1 = 0)}{\mathcal{P}(X_2 = j)} \\
&= \frac{\mathcal{P}(X_1 = 0)\mathcal{P}(X_2 = j \,|\, X_1 = 0)}{\mathcal{P}(X_1 = 0, X_2 = j) + \mathcal{P}(X_1 = 1, X_2 = j)} \\
&= \frac{\mathcal{P}(X_1 = 0)\mathcal{P}(X_2 = j \,|\, X_1 = 0)}{\mathcal{P}(X_1 = 0)\mathcal{P}(X_2 = j \,|\, X_1 = 0) + \mathcal{P}(X_1 = 1)\mathcal{P}(X_2 = j \,|\, X_1 = 1)} \\
\mathcal{P}(X_1 = 0 \,|\, X_2 = 0, X_3 = 0) &= \frac{0.5 \times 0.2}{0.5 \times 0.2 + 0.5 \times 0.3} = 0.4 \\
\mathcal{P}(X_1 = 0 \,|\, X_2 = 2, X_3 = 0) &= \frac{0.5 \times 0.6}{0.5 \times 0.6 + 0.5 \times 0.4} = 0.6
\end{aligned}
$$

因此 $\langle \mathrm{X}_1, \mathrm{X}_2 \,|\, \{\mathrm{X}_3\} \rangle \notin \mathrm{I}(\mathcal{P})$。

最后，对于 $\langle \mathrm{X}_2, \mathrm{X}_3 \,|\, \{\mathrm{X}_1\} \rangle$，考虑任意 $\mathrm{k} \in \{0, 1\}$。利用上面得到的 $\mathcal{P}(\mathrm{X}_3 = 0 \,|\, \mathrm{X}_1 = 0)$，我们有

$$\begin{aligned}\mathcal{P}(X_2 = 0 \mid X_1 = 0, X_3 = k) &= \frac{\mathcal{P}(X_1 = 0, X_2 = 0, X_3 = k)}{\mathcal{P}(X_1 = 0, X_3 = k)} \\ &= \frac{\mathcal{P}(X_1 = 0)\mathcal{P}(X_2 = 0 \mid X_1 = 0)\mathcal{P}(X_3 = k \mid X_2 = 0)}{\mathcal{P}(X_1 = 0)\mathcal{P}(X_3 = k \mid X_1 = 0)} \\ &= \frac{\mathcal{P}(X_2 = 0 \mid X_1 = 0)\mathcal{P}(X_3 = k \mid X_2 = 0)}{\mathcal{P}(X_3 = k \mid X_1 = 0)} \\ \mathcal{P}(X_2 = 0 \mid X_1 = 0, X_3 = 0) &= \frac{0.2 \times 0.2}{0.3} = 0.133 \\ \mathcal{P}(X_2 = 0 \mid X_1 = 0, X_3 = 1) &= \frac{0.2 \times 0.8}{(1 - 0.3)} = 0.229\end{aligned}$$

因此 $\langle X_2, X_3 \mid \{X_1\}\rangle \notin I(\mathcal{P})$。 □

E.2 例 5.7

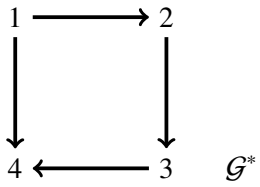


图13： $\mathcal{G}^*$ 来自图 5

已知 V = {X1,X2,X3,X4}，考虑 r(1) = r(2) = r(3) = 2，以及 r(4) = 3。我们希望构造一个多项分布 $\mathcal{P}$，使得 $I(\mathcal{P})$ = {⟨X1, X3, |{X2}⟩, ⟨X2, X4 |{X1, X3}⟩, ⟨X1, X4⟩}。首先，我们从图13的 $\mathcal{G}^*$ 开始，其中 $I(\mathcal{G}^*)$ = {⟨X1, X3, |{X2}⟩, ⟨X2, X4 |{X1, X3}⟩}。下面是 $\boldsymbol{\theta}$ 表。

$\boldsymbol{\theta}_{1,\mathcal{G}^*}$

i	0	1
$\mathcal{P}(X_1 = i)$	a = 0.5	0.5

$\boldsymbol{\theta}_{2,\mathcal{G}^*}$

i	0	1
$\mathcal{P}(X_2 = i \mid X_1 = 0)$	$b_0 = 0.6$	0.4
$\mathcal{P}(X_2 = i \mid X_1 = 1)$	$b_1 = 0.1$	0.9

$\boldsymbol{\theta}_{3,\mathcal{G}^*}$

i	0	1
$\mathcal{P}(X_3 = i \mid X_2 = 0)$	$c_0 = 0.7$	0.3
$\mathcal{P}(X_3 = i \mid X_2 = 1)$	$c_1 = 0.2$	0.8

$\boldsymbol{\theta}_{4,\mathcal{G}^*}$

i	0	1	2
$\mathcal{P}(X_4 = i \mid X_1 = 0, X_3 = 0)$	$d_{00} = 0.1$	$e_{00} = 0.1$	0.8
$\mathcal{P}(X_4 = i \mid X_1 = 0, X_3 = 1)$	$d_{01} = 0.05$	$e_{01} = 0.05$	0.9
$\mathcal{P}(X_4 = i \mid X_1 = 1, X_3 = 0)$	$d_{10} = 0.1125$	$e_{10} = 0.1125$	0.775
$\mathcal{P}(X_4 = i \mid X_1 = 1, X_3 = 1)$	$d_{11} = 0.0625$	$e_{11} = 0.0625$	0.875

要验证 $\langle X_1, X_4\rangle \in I(\mathcal{P})$，只需证明对每个 $l \in \{0, 1\}$，$\mathcal{P}(X_4 = l \mid X_1 = 0) = \mathcal{P}(X_4 = l \mid X_1 = 1)$。考虑每个 $i \in \{0, 1\}$。

$$
\begin{aligned}
\mathcal{P}(X_4 = 0 \mid X_1 = i) &= \sum_{k\in\{0,1\}} \mathcal{P}(X_3 = k, X_4 = 0 \mid X_1 = i) = \sum_{k\in\{0,1\}} \frac{\mathcal{P}(X_1 = i, X_3 = k, X_4 = 0)}{\mathcal{P}(X_1 = i)} \\
&= \sum_{k\in\{0,1\}} \frac{\mathcal{P}(X_4 = 0 \mid X_1 = i, X_3 = k)\mathcal{P}(X_1 = i, X_3 = k)}{\mathcal{P}(X_1 = i)} \\
&= \sum_{k\in\{0,1\}} \frac{\mathcal{P}(X_4 = 0 \mid X_1 = i, X_3 = k)\ \sum_{j\in\{0,1\}} \mathcal{P}(X_1 = i, X_2 = j, X_3 = k)}{\mathcal{P}(X_1 = i)} \\
&= \sum_{k\in\{0,1\}} \left(\frac{\mathcal{P}(X_4 = 0 \mid X_1 = i, X_3 = k)}{\mathcal{P}(X_1 = i)} \sum_{j\in\{0,1\}} \mathcal{P}(X_1 = i)\mathcal{P}(X_2 = j \mid X_1 = i)\mathcal{P}(X_3 = k \mid X_2 = j) \right) \\
&= \sum_{k\in\{0,1\}} \left(\mathcal{P}(X_4 = 0 \mid X_1 = i, X_3 = k) \sum_{j\in\{0,1\}} \mathcal{P}(X_2 = j \mid X_1 = i)\mathcal{P}(X_3 = k \mid X_2 = j) \right) \\
&= d_{i0}(b_i c_0 + (1 - b_i)c_1) + d_{i1}(b_i(1 - c_0) + (1 - b_i)(1 - c_1))
\end{aligned}
$$

$$
\mathcal{P}(X_4 = 0 \mid X_1 = 0) = \Big(0.1 \times (0.6 \times 0.7 + 0.4 \times 0.2)\Big) + \Big(0.05 \times (0.6 \times 0.3 + 0.4 \times 0.8)\Big) = 0.075
$$

$$
\mathcal{P}(X_4 = 0 \mid X_1 = 1) = \Big(0.1125 \times (0.1 \times 0.7 + 0.9 \times 0.2)\Big) + \Big(0.0625 \times (0.1 \times 0.3 + 0.9 \times 0.8)\Big) = 0.075
$$

对于 $\mathcal{P}(\mathrm{X}_4 = 1 \mid \mathrm{X}_1 = 0) = \mathcal{P}(\mathrm{X}_4 = 1 \mid \mathrm{X}_1 = 1)$ 的情形，由于 $\boldsymbol{\theta}_{4,\mathcal{G}^*}$ 的表中有两列相同，因此完全相同。因此，$\langle \mathrm{X}_1, \mathrm{X}_4 \rangle \in \mathrm{I}(\mathcal{P})$。额外条件独立性的检查步骤留给读者。□

E.3 示例 5.9

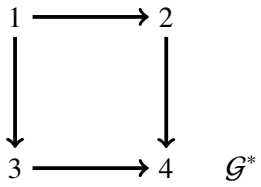


图 14: $\mathcal{G}^*$ 摘自图 6

考虑 $\mathrm{V} = \{\mathrm{X}_1, \mathrm{X}_2, \mathrm{X}_3, \mathrm{X}_4\}$，其中所有变量均为二元变量。我们要构造一个多项分布 $\mathcal{P}$，使得 $\mathrm{I}(\mathcal{P}) = \{\langle \mathrm{X}_1, \mathrm{X}_4 \mid \{\mathrm{X}_2, \mathrm{X}_3\}\rangle, \langle \mathrm{X}_2, \mathrm{X}_3 \mid \{\mathrm{X}_1\}\rangle, \langle \mathrm{X}_1, \mathrm{X}_4\rangle\}$。首先，我们从图 14 中的 $\mathcal{G}^*$ 开始，其中 $\mathrm{I}(\mathcal{G}^*) = \{\langle \mathrm{X}_1, \mathrm{X}_4 \mid \{\mathrm{X}_2, \mathrm{X}_3\}\rangle, \langle \mathrm{X}_2, \mathrm{X}_3 \mid \{\mathrm{X}_1\}\rangle\}$。下面是四个 $\boldsymbol{\theta}$ 表。

$\boldsymbol{\theta}_{1,\mathcal{G}^*}$

i	0	1
$\mathcal{P}(\mathrm{X}_1 = \mathrm{i})$	$\mathrm{a} = 0.5$	0.5

$\boldsymbol{\theta}_{2,\mathcal{G}^*}$

i	0	1
$\mathcal{P}(\mathrm{X}_2 = \mathrm{i} \mid \mathrm{X}_1 = 0)$	$\mathrm{b}_0 = 0.3$	0.7
$\mathcal{P}(\mathrm{X}_2 = \mathrm{i} \mid \mathrm{X}_1 = 1)$	$\mathrm{b}_1 = 0.4$	0.6

$\boldsymbol{\theta}_{3,\mathcal{G}^*}$

i	0	1
$\mathcal{P}(\mathrm{X}_3 = \mathrm{i} \mid \mathrm{X}_1 = 0)$	$\mathrm{c}_0 = 0.4$	0.6
$\mathcal{P}(\mathrm{X}_3 = \mathrm{i} \mid \mathrm{X}_1 = 1)$	$\mathrm{c}_1 = 0.3$	0.7

$\boldsymbol{\theta}_{4,\mathcal{G}^*}$

i	0	1
$\mathcal{P}(\mathrm{X}_4 = \mathrm{i} \mid \mathrm{X}_2 = 0, \mathrm{X}_3 = 0)$	$\mathrm{d}_{00} = 0.1$	0.9
$\mathcal{P}(\mathrm{X}_4 = \mathrm{i} \mid \mathrm{X}_2 = 0, \mathrm{X}_3 = 1)$	$\mathrm{d}_{01} = 0.7$	0.3
$\mathcal{P}(\mathrm{X}_4 = \mathrm{i} \mid \mathrm{X}_2 = 1, \mathrm{X}_3 = 0)$	$\mathrm{d}_{10} = 0.7$	0.3
$\mathcal{P}(\mathrm{X}_4 = \mathrm{i} \mid \mathrm{X}_2 = 1, \mathrm{X}_3 = 1)$	$\mathrm{d}_{11} = 0.8$	0.2

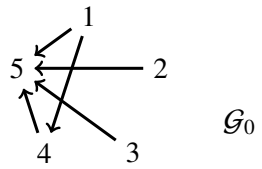
要验证 $\langle \mathrm{X}_1, \mathrm{X}_4 \rangle \in \mathrm{I}(\mathcal{P})$，只需证明 $\mathcal{P}(\mathrm{X}_4 = 0 \,|\, \mathrm{X}_1 = 0) = \mathcal{P}(\mathrm{X}_4 = 0 \,|\, \mathrm{X}_1 = 1)$。任取 $\mathrm{i} \in \{0, 1\}$。

$$
\begin{aligned}
\mathcal{P}(X_4 = 0 \,|\, X_1 = i) &= \sum_{j \in \{0,1\}} \mathcal{P}(X_2 = j, X_4 = 0 \,|\, X_1 = i) \\
&= \sum_{j \in \{0,1\}} \sum_{k \in \{0,1\}} \mathcal{P}(X_2 = j, X_3 = k, X_4 = 0 \,|\, X_1 = i) \\
&= \sum_{j \in \{0,1\}} \sum_{k \in \{0,1\}} \frac{\mathcal{P}(X_1 = i, X_2 = j, X_3 = k, X_4 = 0)}{\mathcal{P}(X_1 = i)} \\
&= \sum_{j \in \{0,1\}} \sum_{k \in \{0,1\}} \frac{\mathcal{P}(X_1 = i)\mathcal{P}(X_2 = j \,|\, X_1 = i)\mathcal{P}(X_3 = k \,|\, X_1 = i)\mathcal{P}(X_4 = 0 \,|\, X_2 = j, X_3 = k)}{\mathcal{P}(X_1 = i)} \\
&= \sum_{j \in \{0,1\}} \mathcal{P}(X_2 = j \,|\, X_1 = i) \sum_{k \in \{0,1\}} \mathcal{P}(X_3 = k \,|\, X_1 = i)\mathcal{P}(X_4 = 0 \,|\, X_2 = j, X_3 = k) \\
&= b_i(c_i d_{00} + (1 - c_i)d_{01}) + (1 - b_i)(c_i d_{10} + (1 - c_i)d_{11}) \\
&= b_i(c_i(d_{00} - d_{01}) + d_{01}) + (1 - b_i)(c_i(d_{10} - d_{11}) + d_{11})
\end{aligned}
$$

$\mathcal{P}(\mathrm{X}_4 = 0 \,|\, \mathrm{X}_1 = 0) = 0.3 \times (0.4 \times (0.1 - 0.7) + 0.7) + (1 - 0.3) \times (0.4 \times (0.7 - 0.8) + 0.8) = 0.67\mathcal{P}($
$\mathrm{X}_4 = 0 \,|\, \mathrm{X}_1 = 1) = 0.4 \times (0.3 \times (0.1 - 0.7) + 0.7) + (1 - 0.4) \times (0.3 \times (0.7 - 0.8) + 0.8) = 0.67$

因此 $\langle \mathrm{X}_1, \mathrm{X}_4 \rangle \in \mathrm{I}(\mathcal{P})$。我们将检查是否存在额外条件独立关系这一步留给读者。 □

E.4 例 5.10



$\mathcal{G}_0$

图15： $\mathcal{G}_0$ 来自图 4

考虑 $\mathrm{V} = \{\mathrm{X}_1, \mathrm{X}_2, \mathrm{X}_3, \mathrm{X}_4, \mathrm{X}_5\}$，其中前四个变量是二元变量，最后一个是三元变量。根据图15中的$\mathcal{G}_0$，下面列出了 $\mathrm{I}(\mathcal{G}_0)$ 中的全部 20 个条件独立关系。

$$
\mathtt{I}(\mathcal{G}_0) = \left\{ \begin{array}{ccc}
\langle X_1, X_2 \rangle, & \langle X_1, X_3 \rangle, & \langle X_2, X_3 \rangle, \\
\langle X_2, X_4 \rangle, & \langle X_3, X_4 \rangle, & \langle X_1, X_2 \,|\, \{X_3\} \rangle, \\
\langle X_1, X_2 \,|\, \{X_4\} \rangle, & \langle X_1, X_3 \,|\, \{X_2\} \rangle, & \langle X_2, X_3 \,|\, \{X_1\} \rangle, \\
\langle X_1, X_3 \,|\, \{X_4\} \rangle, & \langle X_2, X_3 \,|\, \{X_4\} \rangle, & \langle X_2, X_4 \,|\, \{X_1\} \rangle, \\
\langle X_2, X_4 \,|\, \{X_3\} \rangle, & \langle X_3, X_4 \,|\, \{X_1\} \rangle, & \langle X_3, X_4 \,|\, \{X_2\} \rangle, \\
\langle X_1, X_2 \,|\, \{X_3, X_4\} \rangle, & \langle X_1, X_3 \,|\, \{X_2, X_4\} \rangle, & \langle X_2, X_3 \,|\, \{X_1, X_4\} \rangle, \\
\langle X_2, X_4 \,|\, \{X_1, X_3\} \rangle, & \langle X_3, X_4 \,|\, \{X_1, X_2\} \rangle &
\end{array} \right\}
$$

以下是不忠实的条件独立关系集合。

$$
\Psi_{\mathcal{G}_0, \mathring{\mathcal{P}}} = \left\{ \begin{array}{cc}
\psi_1 : \langle X_1, X_5 \rangle, & \psi_2 : \langle X_1, X_2 \,|\, \{X_5\} \rangle, \\
\psi_3 : \langle X_1, X_3 \,|\, \{X_5\} \rangle, & \psi_4 : \langle X_1, X_5 \,|\, \{X_2\} \rangle, \\
\psi_5 : \langle X_1, X_5 \,|\, \{X_3\} \rangle, & \psi_6 : \langle X_1, X_2 \,|\, \{X_3, X_5\} \rangle, \\
\psi_7 : \langle X_1, X_3 \,|\, \{X_2, X_5\} \rangle, & \psi_8 : \langle X_1, X_5 \,|\, \{X_2, X_3\} \rangle
\end{array} \right\}.
$$

与之前只需通过参数诱导出一个条件独立关系的例子不同，我们需要确保满足 $\Psi_{\mathcal{G}_0,\, \mathring{P}}$ 中的八个条件独立关系。以下是五个 $\boldsymbol{\theta}$-表。

$\boldsymbol{\theta}_{1,\mathcal{G}_0}$

i	0	1
$\mathcal{P}(X_1 = i)$	$a = 0.5$	0.5

$\boldsymbol{\theta}_{2,\mathcal{G}_0}$

i	0	1
$\mathcal{P}(X_2 = i)$	$b = 0.5$	0.5

$\boldsymbol{\theta}_{3,\mathcal{G}_0}$

i	0	1
$\mathcal{P}(X_3 = i)$	$c = 0.5$	0.5

$\boldsymbol{\theta}_{4,\mathcal{G}_0}$

i	0	1
$\mathcal{P}(X_4 = i \mid X_1 = 0)$	$d_0 = 0.4$	0.6
$\mathcal{P}(X_4 = i \mid X_1 = 1)$	$d_1 = 0.8$	0.2

$\boldsymbol{\theta}_{5,\mathcal{G}_0}$

i	0	1	2
$\mathcal{P}(X_5 = i \mid X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 0)$	$e_{0000} = 0.4$	$f_{0000} = 0.4$	0.2
$\mathcal{P}(X_5 = i \mid X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 1)$	$e_{0001} = 0.15$	$f_{0001} = 0.15$	0.7
$\mathcal{P}(X_5 = i \mid X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 0)$	$e_{0010} = 0.2$	$f_{0010} = 0.2$	0.6
$\mathcal{P}(X_5 = i \mid X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 1)$	$e_{0011} = 0.2$	$f_{0011} = 0.2$	0.6
$\mathcal{P}(X_5 = i \mid X_1 = 0, X_2 = 1, X_3 = 0, X_4 = 0)$	$e_{0100} = 0.3$	$f_{0100} = 0.3$	0.4
$\mathcal{P}(X_5 = i \mid X_1 = 0, X_2 = 1, X_3 = 0, X_4 = 1)$	$e_{0101} = 0.08$	$f_{0101} = 0.08$	0.84
$\mathcal{P}(X_5 = i \mid X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0)$	$e_{0110} = 0.1$	$f_{0110} = 0.1$	0.8
$\mathcal{P}(X_5 = i \mid X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 1)$	$e_{0111} = 0.3$	$f_{0111} = 0.3$	0.4
$\mathcal{P}(X_5 = i \mid X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 0)$	$e_{1000} = 0.2$	$f_{1000} = 0.2$	0.6
$\mathcal{P}(X_5 = i \mid X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 1)$	$e_{1001} = 0.45$	$f_{1001} = 0.45$	0.1
$\mathcal{P}(X_5 = i \mid X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 0)$	$e_{1010} = 0.15$	$f_{1010} = 0.15$	0.7
$\mathcal{P}(X_5 = i \mid X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1)$	$e_{1011} = 0.4$	$f_{1011} = 0.4$	0.2
$\mathcal{P}(X_5 = i \mid X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 0)$	$e_{1100} = 0.15$	$f_{1100} = 0.15$	0.7
$\mathcal{P}(X_5 = i \mid X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 1)$	$e_{1101} = 0.24$	$f_{1101} = 0.24$	0.52
$\mathcal{P}(X_5 = i \mid X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 0)$	$e_{1110} = 0.225$	$f_{1110} = 0.225$	0.55
$\mathcal{P}(X_5 = i \mid X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 1)$	$e_{1111} = 0.2$	$f_{1111} = 0.2$	0.6

特别考虑 $\psi_8 = \langle X_1, X_5 \mid \{X_2, X_3\}\rangle$。我们首先要证明，如果 $\psi_8 \in \mathrm{I}(\mathcal{P})$，那么 $\Psi_{\mathcal{G}_0,\mathring{P}}$ 中的其他七个不忠实条件独立也都属于 $\mathrm{I}(\mathcal{P})$。为此，注意到 $\mathcal{P}$ 是严格正的多项分布，因为五个 $\boldsymbol{\theta}$-表中的每个单元格都为正。因此， $\mathcal{P}$是一个拟阵。现在我们证明， ψ_1， ...， ψ_7 中的每一项都可以利用附录 A 中讨论的拟阵独立公理（包括交集）从 $\mathrm{I}(\mathcal{G}_0) \cup \{\psi_8\}$ 推导出来。具体而言，记 $\phi_1 = \langle X_1, X_2\rangle$、$\phi_2 = \langle X_1, X_2 \mid \{X_3\}\rangle$和 $\phi_3 = \langle X_1, X_3 \mid \{X_2\}\rangle$，其中 ϕ_1、ϕ_2、$\phi_3 \in \mathrm{I}(\mathcal{G}_0) \subset \mathrm{I}(\mathcal{P})$。

$$\underbrace{(X_1 \perp\!\!\!\perp_{\mathcal{P}} X_2 \,|\, \{X_3\})}_{\phi_2} \wedge \underbrace{(X_1 \perp\!\!\!\perp_{\mathcal{P}} X_5 \,|\, \{X_2, X_3\})}_{\psi_8} \Rightarrow \underbrace{X_1 \perp\!\!\!\perp_{\mathcal{P}} \{X_2, X_5\} \,|\, \{X_3\}}_{\text{(i)}} \quad \text{收缩} \tag{13}$$

$$(\mathrm{X}_1 \perp\!\!\!\perp_{\mathcal{P}} \underbrace{\mathrm{X}_2}_{\phi_2} \,|\, \{\mathrm{X}_3\}) \wedge (\mathrm{X}_1 \perp\!\!\!\perp_{\mathcal{P}} \underbrace{\mathrm{X}_5}_{\psi_8} \,|\, \{\mathrm{X}_2, \mathrm{X}_3\}) \Rightarrow \mathrm{X}_1 \perp\!\!\!\perp_{\mathcal{P}} \{\underbrace{\mathrm{X}_2, \mathrm{X}_5}_{\text{(i)}}\} \,|\, \{\mathrm{X}_3\} \quad \text{收缩} \tag{14}$$

$$\Rightarrow \underbrace{X_1 \perp\!\!\!\perp_{\mathcal{P}} X_2 \,|\, \{X_3, X_5\}}_{\psi_6} \quad \text{(i)，弱并} \tag{15}$$

$$\Rightarrow \underbrace{X_1 \perp\!\!\!\perp_{\mathcal{P}} X_5 \,|\, \{X_3\}}_{\psi_5} \quad \text{(i)，分解} \tag{16}$$

$$(\mathrm{X}_1 \perp\!\!\!\perp_{\mathcal{P}} \underbrace{\mathrm{X}_3}_{\phi_3} \,|\, \{\mathrm{X}_2\}) \wedge (\mathrm{X}_1 \perp\!\!\!\perp_{\mathcal{P}} \underbrace{\mathrm{X}_5}_{\psi_8} \,|\, \{\mathrm{X}_2, \mathrm{X}_3\}) \Rightarrow \mathrm{X}_1 \perp\!\!\!\perp_{\mathcal{P}} \{\underbrace{\mathrm{X}_3, \mathrm{X}_5}_{\text{(ii)}}\} \,|\, \{\mathrm{X}_2\} \quad \text{收缩} \tag{17}$$

$$\Rightarrow \underbrace{X_1 \perp\!\!\!\perp_{\mathcal{P}} X_3 \,|\, \{X_2, X_5\}}_{\psi_7} \quad \text{(ii)，弱并} \tag{18}$$

$$\Rightarrow \underbrace{X_1 \perp\!\!\!\perp_{\mathcal{P}} X_5 \,|\, \{X_2\}}_{\psi_4} \quad \text{(ii)，分解} \tag{19}$$

$$\mathrm{X}_1 \perp\!\!\!\perp_{\mathcal{P}} \underbrace{\mathrm{X}_2}_{\psi_6} \,|\, \{\mathrm{X}_3, \mathrm{X}_5\} \wedge \mathrm{X}_1 \perp\!\!\!\perp_{\mathcal{P}} \underbrace{\mathrm{X}_3}_{\psi_7} \,|\, \{\mathrm{X}_2, \mathrm{X}_5\} \Rightarrow \mathrm{X}_1 \perp\!\!\!\perp_{\mathcal{P}} \{\underbrace{\mathrm{X}_2, \mathrm{X}_3}_{\text{(iii)}}\} \,|\, \{\mathrm{X}_5\} \quad \text{交集} \tag{20}$$

$$\Rightarrow \underbrace{X_1 \perp\!\!\!\perp_{\mathcal{P}} X_2 \,|\, \{X_5\}}_{\psi_2} \quad \text{(iii)，分解} \tag{21}$$

$$\Rightarrow \underbrace{X_1 \perp\!\!\!\perp_{\mathcal{P}} X_3 \,|\, \{X_5\}}_{\psi_3} \quad \text{(iii)，分解} \tag{22}$$

$$\underbrace{(X_1 \perp\!\!\!\perp_{\mathcal{P}} X_2)}_{\phi_1} \wedge \underbrace{(X_1, X_5 \,|\, \{X_2\})}_{\psi_4} \Rightarrow \underbrace{X_1 \perp\!\!\!\perp_{\mathcal{P}} \{X_2, X_5\}}_{\text{(iv)}} \quad \text{收缩} \tag{23}$$

$$\Rightarrow \underbrace{X_1 \perp\!\!\!\perp_{\mathcal{P}} X_5}_{\psi_1} \quad \text{(iv)，分解} \tag{24}$$

因此，只需证明 $\psi_8 \in \mathrm{I}(\mathcal{P})$，也就是说，对于每个 j、$\mathrm{k} \in \{0,1\}$，都有 $\mathcal{P}(\mathrm{X}_5 = 0 \,|\, \mathrm{X}_1 = 0, \mathrm{X}_2 = \mathrm{j}, \mathrm{X}_3 = \mathrm{k}) = \mathcal{P}(\mathrm{X}_5 = 0 \,|\, \mathrm{X}_1 = 1, \mathrm{X}_2 = \mathrm{j}, \mathrm{X}_3 = \mathrm{k})$。任取 i、j、$\mathrm{k} \in \{0,1\}$。

$$\begin{aligned}
&\mathcal{P}(X_5 = 0 \,|\, X_1 = i, X_2 = j, X_3 = k) \\
&= \sum_{l \in \{0,1\}} \mathcal{P}(X_4 = l, X_5 = 0 \,|\, X_1 = i, X_2 = j, X_3 = k) \\
&= \sum_{l \in \{0,1\}} \frac{\mathcal{P}(X_1 = i)\mathcal{P}(X_2 = j)\mathcal{P}(X_3 = k)\mathcal{P}(X_4 = l \,|\, X_1 = i)\mathcal{P}(X_5 = 0 \,|\, X_1 = i, X_2 = j, X_3 = k, X_4 = l)}{\mathcal{P}(X_1 = i)\mathcal{P}(X_2 = j)\mathcal{P}(X_3 = k)} \\
&= \sum_{l \in \{0,1\}} \mathcal{P}(X_4 = l \,|\, X_1 = i)\mathcal{P}(X_5 = 0 \,|\, X_1 = i, X_2 = j, X_3 = k, X_4 = l) \\
&= d_i e_{ijk0} + (1 - d_i) e_{ijk1}
\end{aligned}$$

因此，我们有：

$$\begin{aligned}
\mathcal{P}(X_5 = 0 \,|\, X_1 = 0, X_2 = j, X_3 = k) &= 0.4 e_{0jk0} + 0.6 e_{0jk1} \\
\mathcal{P}(X_5 = 0 \,|\, X_1 = 1, X_2 = j, X_3 = k) &= 0.8 e_{1jk0} + 0.2 e_{1jk1}
\end{aligned}$$

$$\mathcal{P}(X_5 = 0 \mid X_1 = 0, X_2 = 0, X_3 = 0) = 0.4 \times 0.4 + 0.6 \times 0.15 = 0.25$$
$$\mathcal{P}(X_5 = 0 \mid X_1 = 1, X_2 = 0, X_3 = 0) = 0.8 \times 0.2 + 0.2 \times 0.45 = 0.25$$
$$\mathcal{P}(X_5 = 0 \mid X_1 = 0, X_2 = 0, X_3 = 1) = 0.4 \times 0.2 + 0.6 \times 0.2 = 0.2$$
$$\mathcal{P}(X_5 = 0 \mid X_1 = 1, X_2 = 0, X_3 = 1) = 0.8 \times 0.15 + 0.2 \times 0.4 = 0.2$$
$$\mathcal{P}(X_5 = 0 \mid X_1 = 0, X_2 = 1, X_3 = 0) = 0.4 \times 0.3 + 0.6 \times 0.08 = 0.168$$
$$\mathcal{P}(X_5 = 0 \mid X_1 = 1, X_2 = 1, X_3 = 0) = 0.8 \times 0.15 + 0.2 \times 0.24 = 0.168$$
$$\mathcal{P}(X_5 = 0 \mid X_1 = 0, X_2 = 1, X_3 = 1) = 0.4 \times 0.1 + 0.6 \times 0.3 = 0.22$$
$$\mathcal{P}(X_5 = 0 \mid X_1 = 1, X_2 = 1, X_3 = 1) = 0.8 \times 0.225 + 0.2 \times 0.2 = 0.22$$

对于每个 j,k，$\mathcal{P}(X_5 = 1 \mid X_1 = 0, X_2 = j, X_3 = k) = \mathcal{P}(X_5 = 1 \mid X_1 = 1, X_2 = j, X_3 = k)$ 的情形完全相同，因为 $\theta_{5,\ \mathcal{G}_0}$ 的表中有两列相同。因此，我们有$\psi_8 = \langle X_1, X_5 \mid \{X_2, X_3\}\rangle \in I(\mathcal{P})$。