arXiv:2304.01953v1 [stat.ME] 4 Apr 2023

纠缠缺失的图模型

BY RANJANI SRINIVASAN[1,a], ROHIT BHATTACHARYA[2,b], RAZIEH NABI[3,c], ELIZABETH L. OGBURN[4,d], AND ILYA SHPITSER[5,e]

[1]约翰斯·霍普金斯大学电气与计算机工程系，a*ranjani@jhu.edu*

[2]威廉姆斯学院计算机科学系，b*rb17@williams.edu*

[3]埃默里大学生物统计学与生物信息学系,c*razieh.nabi@emory.edu*

[4]约翰·霍普金斯大学生物统计学系，d*eogburn@jhsph.edu*

[5]约翰斯·霍普金斯大学计算机科学系，伊*ilyas@cs.jhu.edu*

尽管针对存在数据依赖的情形进行因果推断和统计推断日益受到关注，但目前用于处理依赖数据情形下缺失数据的方法仍然寥寥无几；统计学和因果推断领域的大多数经典缺失数据方法都将数据单元视为独立同分布（i.i.d.）。我们提出一种基于图模型的框架，用于在存在纠缠缺失时进行因果推断，其中纠缠缺失定义为存在数据依赖的缺失性。我们区分了可能出现的三种不同类型的纠缠，并以真实世界的实例加以说明。对于这三种情形，我们均给出了可靠且完备的识别结果。我们表明，现有的缺失数据模型可以扩展，以涵盖由（1）目标分布依赖和（2）缺失过程依赖引起的纠缠；而由（3）缺失干扰引起的纠缠则需要采用一种新方法。我们通过合成数据展示了纠缠缺失框架的用法。最后，我们讨论了在对模型中的变量作出特定重新解释的前提下，我们的缺失干扰模型如何将缺失数据方法扩展到独立同分布设置中的新型缺失数据模式。

1. 引言。 缺失数据是一个普遍存在的问题：数据可能由于调查无应答、脱落或失访、不完善的数据收集，或其他复杂因素而缺失。如果未得到适当处理，系统性缺失的数据记录可能会使后续分析产生显著偏差。大多数关于缺失数据的文献将数据个体视为独立同分布（i.i.d.）[1–16]。在许多类型的问题中，这一假设是合理的，因为个体之间的相互作用可以忽略不计。然而，近年来人们越来越认识到，在许多情境下，数据在个体之间存在依赖，尤其会受到干扰的影响，即在一个个体上测量的变量可能会对另一个个体上测量的变量产生因果效应。典型例子是传染病情境，在这种情境下，对一个个体进行治疗可能会对其他个体产生保护作用；但近年来，人们开始关注人类通过其社会网络（广义上包括家庭、工作、社交、邻域或在线关系）以各种方式相互影响的问题 [17–45]。

现有文献对数据依赖与缺失数据可能同时出现这一情况认识不足，只有少数研究例外： [46] 研究了健康数据网络中缺失数据的多重插补技术； [47] 实证描述了系统性缺失数据下网络数据分析中的偏差； [48] 采用了基于似然的

针对部分观测数据的健康研究建模方法； [49] 提出了存在缺失数据时网络逻辑回归模型的估计方法。

据我们所知，本文提出了首个用于处理纠缠缺失情形的非参数[1] 方法，即同时存在缺失性和数据依赖的情形；本文也是首个基于因果图模型处理此类情形的方法。我们开发了一个图建模框架，用于处理三种不同类型的纠缠缺失，分别称为目标分布依赖、缺失过程依赖和缺失干扰。我们针对缺失性与数据依赖可能相互作用的上述三种情形，推导出了可靠且完备的非参数识别结果。我们的结果为存在纠缠缺失时的因果估计目标提供了完全一般性的处理。

目标分布依赖是指完整数据分布存在统计依赖或因果依赖。例如，考虑一项关于疫苗有效性的研究，其中一个个体的疫苗接种状态可能有助于保护其朋友或家庭成员免于感染传染病。潜在的完整数据分布是非独立同分布的；在插补缺失数据或以其他方式对其进行控制时，若未考虑这种依赖，通常会引入偏差。

缺失过程依赖 是指一个单位的缺失状态可能取决于其他单位的变量。例如，个人是否参与调查，可能取决于其朋友或家庭成员的决定。类似地，设想一项大规模公共卫生监测工作：研究人员收集大量人群的人口统计和家庭数据，并邀请其中一小部分人参加子研究，以收集更详细的数据。个人是否选择加入子研究，可能既取决于其自身特征，也取决于其社会网络中其他人的特征或选择。

缺失干扰 是指某个变量存在多个潜在结果，而这些结果取决于网络中其他单位的结局是否被测量（对应于这些其他单位是否参与研究）。以一项活动或项目的评估为例，例如通过向参与者发放调查问卷来进行课程评价。对于未参加课程或在课程结束前退出的个人，其调查问题的答案显然无法观测。然而，由于课堂具有社交和协作属性，某个人的回答潜在值——如果该人参加了课程，可能会因其同伴是否出席而有所不同。学生表现评估也有类似情况：有证据表明，课堂构成会影响个别学生的表现 [50]。此前，缺失数据或干扰相关文献尚未探讨这些情形。我们提出，以一种在结构上类似于因果推断中的反事实变量或潜在结果的方式，对潜在满意度或成就变量的多个版本进行建模。

本文结构如下。第 2 节回顾因果图模型、干扰以及当前用于缺失数据的图模型等相关背景。2。第 3 节介绍用于表示纠缠缺失的新图模型。3。第 4 节给出我们关于纠缠缺失的新识别结果。4。第5节表明，我们的缺失干扰模型将缺失数据方法扩展到了独立同分布设定下的新型缺失数据模式。第6 节通过模拟展示了我们框架的实用性。最后，我们在第7节概述了在存在纠缠缺失的设定下，对任意估计目标进行非参数识别仍面临的开放性挑战。

[1]在这里，我们使用“非参数”一词，是指我们不对变量之间关系的泛函形式施加任何限制。

2. 因果图模型、干扰与缺失数据。因果推断传统上关注诸如 $Y(a)$ 这样的反事实或潜在结果，其定义为：如果某个个体接受了处理或暴露值 $A = a$，那么可能与事实相反，该个体本会观察到的结局。为简单起见，我们假设全文中的处理均为二元处理，但我们的结果也可推广到多水平处理，以及（还需考虑一些额外的细微问题）连续型 A。因果效应通常定义为反事实结果之间的对比，例如 $\beta := \mathbb{E}[Y(1) - Y(0)]$，它比较了所有个体都接受处理 $A = 1$ 时反事实结果的期望值，与所有个体都接受处理 $A = 0$ 时反事实结果的期望值。由于潜在结果未必会在可用的数据实现中被观测到，因此需要作出假设，将反事实分布与观测数据分布联系起来；可用样本正是从该分布中抽取的。

用于将 β 表示为观测数据的泛函的一组标准假设包括：(i) 一致性，即观测结果的实现值等于处理设为实际观测值时的反事实结果实现值，或 $Y(a) = Y$ 当 $A = a$ 时，（ii）正值性，即所有处理分配（可能以一组基线协变量 C 为条件）都具有正支持，或者若 $p(\mathsf{C} = \mathsf{c}) > 0$ 则 $p(a \mid \mathsf{C}{=}\mathsf{c}) > 0$ 对 a 在 A 的支持集中的所有取值均成立，以及 (iii) 条件可忽略性，即潜在结果与处理分配相互独立（可能以一组基线协变量 C 为条件），或对所有 a，$Y(a) \perp\!\!\!\perp A \mid \mathsf{C}$。在这些假设下，完整数据分布 $p(Y(1), Y(0), A, \mathsf{C})$ 的参数 β 可通过调整泛函：$\mathbb{E}[\mathbb{E}[Y|A=1,\mathsf{C}] - \mathbb{E}[Y|A=0,\mathsf{C}]]$，由观测数据分布 $p(Y, A, \mathsf{C})$ 识别。

条件独立性假设（如条件可忽略性）可以用图模型表示；本文将采用这种方法来呈现纠缠缺失情况下的识别结果。我们简要介绍本研究中将用到的图形模型记法、术语和概念。我们尽量在本节中涵盖所有必需的概念；如需更全面地了解图模型，请参阅 [51–53]。

2.1. 因果图模型。有向无环图（DAG）$\mathcal{G}$ 的顶点集为 V，表示变量 V；它仅包含有向边（$\to$），且不含有向环。DAG $\mathcal{G}(\mathsf{V})$ 的统计模型是分布 $p(\mathsf{V})$ 的集合，这些分布可分解为 $p(\mathsf{V}) = \prod_{V \in \mathsf{V}} p(V \mid \mathrm{pa}_{\mathcal{G}}(V))$，其中 $\mathrm{pa}_{\mathcal{G}}(V)$ 表示 V 在 $\mathcal{G}(\mathsf{V})$ 中的父节点；若 $U \to V$，则 U 是 V 的父节点，其中 $U, V \in \mathsf{V}$。$p(\mathsf{V})$ 中由 DAG $\mathcal{G}(\mathsf{V})$ 编码的所有条件独立性约束，都可以通过运用 d-分离规则 [54]检查 $\mathcal{G}(\mathsf{V})$ 中的路径来读出。

DAG 的因果模型将 $p(\mathsf{V})$ 的生成模型与干预理论相结合，从而得到反事实随机变量上的分布。这些模型通常使用结构方程来定义（结构方程是不变因果机制的数学表述），具体如下。在因果模型中，每个变量 V 都由其父节点 $\mathrm{pa}_{\mathcal{G}}(V)$ 的取值和一个外生噪声变量 ϵ_V 通过结构方程 $f_V(\mathrm{pa}_{\mathcal{G}}(V), \epsilon_V)$ 决定。给定对 $V \in \mathsf{V}$ 的干预，将 V 设为 v，结构方程 $f_V(\mathrm{pa}_{\mathcal{G}}(V), \epsilon_V)$ 将被替换为输出常数值 v 的方程。结构方程和噪声变量共同产生观测数据分布 $p(\mathsf{V})$。如果将 $\mathsf{A} \subseteq \mathsf{V}$ 的任意子集设为取值 a，则 $p(\mathsf{Y}(\mathsf{a}))$ 表示关于该干预的潜在结果 $\mathsf{Y} = \mathsf{V} \setminus \mathsf{A}$ 的联合分布，在 [51] 中记作 $p(\mathsf{Y} \mid \mathrm{do}(\mathsf{a}))$。就本文而言，我们假设结构方程和噪声项满足以下条件：所得的 $p(\mathsf{V})$ 是一个正分布。

一种常见的因果模型称为结构因果模型，或称为独立误差非参数结构方程模型 *(NPSEM-IE)* [55]，它假设联合分布

所有外生项在边际上相互独立： $p(\epsilon) = \prod_{V \in \mathbf{V}} p(\epsilon_V)$。NPSEM-IE蕴含相对于 $\mathcal{G}(\mathbf{V})$的 $p(\mathbf{V})$ DAG因子分解，以及一种称为*g*-公式的修正DAG因子分解。该公式将每个联合干预分布识别为观测数据分布 $p(\mathbf{V})$的如下泛函：

$$p(\mathbf{Y}(\mathbf{a})) = \prod_{Y \in \mathbf{Y}} p(Y \mid \mathrm{pa}_{\mathcal{G}}(Y))\Big|_{\mathbf{A}=\mathbf{a}}, \tag{1}$$

对于每个$\mathbf{A} \subset \mathbf{V}$ 和$\mathbf{Y} = \mathbf{V} \setminus \mathbf{A}$。

前文描述的条件可忽略模型可通过图 1(a) 所示的有向无环图表示。对于式 (1) 中的 $\mathbf{V} = \{Y, A, C\}$、$\mathbf{A} = \{A\}$ 和 $\mathbf{V} \setminus \mathbf{A} = \{Y, C\}$， $p(Y(a)\ C(a)) = p(Y(a)\ C) = p(Y \mid A = a, C) \times p(C)$。因此，反事实均值 $\mathbb{E}[Y(a)]$可通过调整公式识别为 $\mathbb{E}[Y(a)] = \sum_C \mathbb{E}[Y \mid A = a, C] \times p(C)$。[2] 等价地，干预目标也可通过逆概率加权（IPW）泛函 $\mathbb{E}[\{\mathbb{I}(A = a)/p(A \mid C)\} \times Y]$ 获得，该泛函涉及使用倾向评分 $p(A \mid C)$ 权重 [56]。在考察缺失数据模型中的识别时，加权方法将尤其有用；在该泛函中，缺失机制将类似于倾向评分（见第 2.3 节）。

含有隐变量的因果模型可能对应于一个有向无环图 $\mathcal{G}(\mathbf{V} \cup \mathbf{H})$，其中$\mathbf{V}$是观测变量，$\mathbf{H}$ 是未观测变量。在这类模型中，并非所有因果效应都能在标准假设下得到识别；识别理论也可以在潜在投影图 $\mathcal{G}(\mathbf{V})$ [57], 上表述，该图仅包含与观测变量相对应的顶点。潜在投影是一种称为无环有向混合图（ADMG）的混合图，包含表示直接因果关系的有向边，以及表示未观测混杂的双向边（$\leftrightarrow$）。例如，若图1(a)中的有向无环图里， C 是未观测变量，则其潜在投影如图1(b)所示。与祖先有向混合图 $\mathcal{G}(\mathbf{V})$ 相关的分布 $p(\mathbf{V})$ 中的条件独立性陈述，可以通过一种称为*m*-分离 [58]的 d-分离推广形式读出。

隐变量模型中可识别的因果效应由ID算法给出。可以将其视为g-公式的一种推广，因为ID算法所识别的因果效应是祖先有向混合图的修正因子分解，正如g-公式是有向无环图的修正因子分解一样。关于由有向无环图和祖先有向混合图表示的因果模型中平均因果效应的一般识别结果，参见 [59–63]。

2.2. 干扰。 传统上，因果推断方法假设所有个体在因果和统计意义上相互独立。然而，这一假设经常不成立；当一个个体可能对其他个体产生因果影响时，我们称存在干扰 [17–45]。有时，干扰被更狭义地定义为一个个体所接受的处理对另一个个体结局的因果效应；我们对这一术语的用法更为宽泛，也适用于协变量、缺失指示变量、处理或结局。

本文为将非参数识别问题与统计依赖问题区分开来，将假设存在部分干扰：干扰可能发生在由固定有限数量单位组成的集合或区块内部，但各区块之间相互独立。此外，我们还将假设各区块同分布。这大大简化了我们的论述，但原则上，我们所有的识别结果都可以推广到完全干扰，或推广到区块

[2]在缺失数据领域及部分因果推断文献中，通常将期望视为关注目标；但为了更好地与因果图模型识别文献衔接，本文将讨论分布的识别。

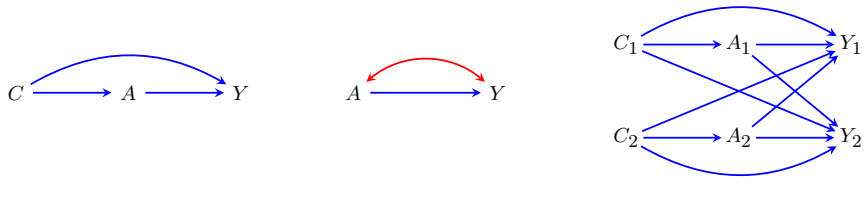


图 1: （a）条件可忽略模型中有向无环图（DAG）表示的示例；（b）若 C 是未观测变量，则对（a）中有向无环图进行潜在投影的结果；（c）二元组中干扰的有向无环图示意图。

结构因不同实现而异的情形。如果某个完整数据估计目标（例如完整数据分布）可以表示为观测数据分布的一个泛函，我们就称其已识别；这一概念不受部分干扰还是完全干扰的影响。另一个相关但独立的问题是统计可估计性，它通常要求目标可识别，且存在大量独立（或弱依赖）观测；非独立同分布的区块可能会给可估计性带来挑战。尽管本文关注的是识别而非估计，但我们指出，在固定大小的独立同分布区块这一设定下，可以采用针对独立同分布数据的标准技术推导估计量，只需在估计时将每个单位块视为一个个体即可。在区块数量趋于无穷、而区块构成保持固定且独立同分布的渐近框架下，传统估计结果将直接成立。这意味着，我们可以选择将每个区块视为由多个单位组成的集合，也可以将其视为一个多变量观测；全文将交替采用这两种视角。

设有 b 个区块，每个区块包含 $m = n/b$ 个单位，其中 n 是单位总数。令 A $\equiv(A_1, \ldots,$ $A_m)$ 表示区块中单位 $i = 1, \ldots m$ 的处理分配向量，m 维向量 a 表示 A 支持集中的一个取值。类似地，令 Y $\equiv(Y_1, \ldots, Y_m)$ 和 C $\equiv$(C1, . . . , C$_m$) 分别表示结局向量和协变量向量。由于同一区块内的单位可能相互作用，因此，对单位 i 的反事实结果进行索引时，不仅要考虑对单位 i 实施的干预，也要考虑对同一区块内其他单位实施的干预。具体而言，我们将 Y_i(a) 定义为单位 i 的反事实结果，其中对处理向量 A 实施干预并将其设为 a。将此与式（1）中的反事实 $Y(a)$ 进行比较：在无干扰的情形下，单位的结局也是对将 A 设为 a 的干预的响应，但此处 A 仅表示该单位自身的处理。

按网络中每个个体的处理进行索引的反事实，可用于定义单个个体的因果效应，或网络中所有个体的平均效应。干扰问题中的感兴趣的参数及相关估计策略详见 [64–67]。

Ogburn 和 VanderWeele [64] 将有向无环图的因果模型从独立同分布设定扩展到了干扰问题。图 1(c) 展示了有向无环图如何表示干扰的典型示例；此处的区块包含 2 个个体（二元组）。在该有向无环图中，元组(C_1, A_1, Y_1)对应个体1 的变量，而元组(C_2, A_2, Y_2) 对应个体2的变量。这两个元组之间存在边，表示一个个体的变量对另一个个体变量的因果影响，即存在干扰。

图形因果模型的原理可从独立同分布设定推广到存在干扰的设定，唯一的区别是，此时图表示的是一个完整的单位块，而不是代表独立同分布实现的单个单位。如上所述，这只是视角切换的问题：我们可以将单位集合视为整体，也可以将单个单位的

多变量观测视为整体。此外，我们能够将基于图的识别算法从独立同分布设置扩展到部分干扰设置。

2.3. 缺失数据模型的图形表示。 缺失数据模型编码了关于缺失性如何产生以及如何与感兴趣的潜在变量相关联的假设。该模型是一组定义在一组随机变量 $\{\mathbf{O}、\mathbf{Z}^{(1)}、\mathbf{R}、\mathbf{Z}\}$ 上的分布，其中 $\mathbf{O}$ 表示始终可观测的变量集合，$\mathbf{Z}^{(1)}$ 表示可能缺失的变量集合，$\mathbf{R}$ 表示$\mathbf{Z}^{(1)}$中各变量的缺失指示变量集合，而$\mathbf{Z}$ 表示观测代理变量集合，这些代理变量对应于$\mathbf{Z}^{(1)}$中的变量。与 $Z^{(1)} \in \mathbf{Z}^{(1)}$ 对应的缺失指示变量$R_Z \in \mathbf{R}$，在变量 Z 可观测时取值为 1，不可观测时取值为 0。观测代理变量 Z 按如下方式确定性定义：

$$Z = \begin{cases} Z^{(1)} & \text{if } R_Z = 1, \\ ? & \text{if } R_Z = 0. \end{cases} \tag{2}$$

按照因果模型的术语，每个缺失指示变量 $R \in \mathbf{R}$ 都可以视为可进行干预的处理变量，而每个 $Z \in \mathbf{Z}$ 都可以视为观测结果。因此， $Z^{(1)}$ 是一个反事实随机变量，表示我们假设已对相应的缺失指示变量进行干预并将其设为 R_Z 1时的情形，尽管事实上我们可能并未这样做。因此，我们可以将（2）视为因果推断中一致性假设在缺失数据情境下的对应形式。

我们将分布 p（$\mathbf{O}$，$\mathbf{Z}^{(1)}$，$\mathbf{R}$）称为完整联合分布，将分布 p（$\mathbf{O}$，$\mathbf{Z}^{(1)}$）称为目标分布，将分布 p（$\mathbf{O}$，$\mathbf{Z}$，$\mathbf{R}$）称为观测数据分布，并将条件分布 p（$\mathbf{R} \mid \mathbf{Z}^{(1)}$，$\mathbf{O}$）称为缺失过程，或缺失数据倾向评分。
通常，我们关注目标分布的某个泛函，其形式通常为 h（$\mathbf{O}$，$\mathbf{Z}^{(1)}$），而缺失数据方法的目标是根据观测数据分布对这类泛函进行非参数识别。这需要作出假设，通常会假设缺失数据的一致性假设（2）和正值性假设的缺失数据版本：

$$p(\mathbf{R} = 1 \mid \mathbf{Z}^{(1)}, \mathbf{O}) > 0. \tag{3}$$

与因果推断中的情况一样，一致性和正值性不足以非参数识别目标分布的参数，但加入某些独立性假设后，可能足以非参数识别完整联合分布，进而识别其任意泛函。当前文献研究的缺失数据模型中，这类假设可编码在有向无环图 $\mathcal{G}$(V) 上，其中顶点 V 对应于 O $\cup$Z$^{(1)}$ $\cup$ R $\cup$Z 中的随机变量，并对 $\mathcal{G}$(V) 施加某些额外限制：（i）每个 $Z \in$Z 只有两个父节点，R_Z 和 $Z^{(1)}$；（ii）R 中的变量不能指向 O $\cup$ Z$^{(1)}$ 中的变量。根据定义，限制（i）通过（2）施加于 Z。为区分（2）所蕴含的确定性关系与概率关系，我们用灰色绘制指向 $Z \in$Z 的边。施加限制（ii）是为了确保：尽管 R_Z 的变化会导致观测代理变量 $Z \in$Z 发生变化，却不会导致底层完整数据 $Z^{(1)} \in$Z$^{(1)}$ 发生变化。我们在以下各节阐述的所提出的框架放宽了限制（ii）中的边限制；更多详情见第 5 节。

缺失数据 DAG 模型是定义在变量 $\mathbf{O} \cup \mathbf{Z}^{(1)} \cup \mathbf{R} \cup \mathbf{Z}$上的一组分布，这些分布相对于满足上述限制的有向无环图进行因子分解，具体如下：

$$p(\mathbf{O}, \mathbf{Z}^{(1)}, \mathbf{R}, \mathbf{Z}) = \prod_{V \in \mathbf{O} \cup \mathbf{Z}^{(1)} \cup \mathbf{R}} p(V \mid \mathrm{pa}_{\mathcal{G}}(V)) \times \prod_{Z \in \mathbf{Z}} p(Z \mid R_Z, Z^{(1)}). \tag{4}$$

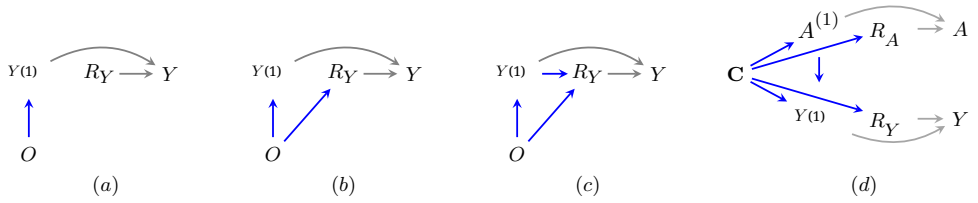


图 2：与（a）完全随机缺失模型、（b）随机缺失模型和（c）非随机缺失模型相对应的缺失数据 DAG 示例。（d）一个带有 $Z \equiv \{A, Y\}$ 的缺失数据随机缺失模型。

缺失数据有向无环图的示例如图 2 所示：图 2(a) 对应于一个 完全随机缺失（MCAR）模型，其中 $R_Y \perp\!\!\!\perp O$，$Y_{(1)}$；图 2(b) 对应于一个随机缺失（MAR）模型，其中 $R_Y \perp\!\!\!\perp Y^{(1)} \mid O$；图 2(c) 对应于一个非随机缺失模型，其中两种独立性均不成立。

如果缺失数据模型包含隐变量 $\mathbf{H}$，则可像隐变量因果模型那样，用 DAG $\mathcal{G}(\mathbf{O},\mathbf{Z},\mathbf{Z}^{(1)},\mathbf{R})$ 的潜在投影无向混合图 $\mathcal{G}(\mathbf{H},\mathbf{O},\mathbf{Z},\mathbf{Z}^{(1)},\mathbf{R})$来表示。在这类模型中，推断目标是完整数据分布 $p(\mathbf{O},\mathbf{Z}^{(1)},\mathbf{R})$ 的某个函数，其中 $\mathbf{H}$ 已被边缘化。

关于如何处理缺失数据，已有大量非图形方法的研究。一种简单的方法是在完全案例分析中仅使用完全观测行，但只有当底层机制为完全随机缺失（MCAR） [1]时，这种做法才有依据。对于随机缺失（MAR）情形，人们已开发出多种方法，包括期望最大化 [3, 68, 69], 、多重插补 [70, 71], 和逆概率加权 [72, 73] 。对于非随机缺失（MNAR）问题，通常会对底层数据分布和缺失选择模型施加参数或半参数限制，使其能够实现识别 [3, 74–79]。在本研究中，我们使用图模型推导非参数识别的条件。

当且仅当缺失机制可由观测数据分布的某个泛函识别时，缺失数据模型中的完整联合分布才可识别当且仅当缺失机制可由观测数据分布 $g(p(\mathbf{R},\ \mathbf{O},\ \mathbf{Z}))$的某个泛函识别，因为

$$(5)\quad p(\mathbf{O},\mathbf{Z}^{(1)},\mathbf{R}) = p(\mathbf{O},\mathbf{Z}^{(1)}) \times \underbrace{p(\mathbf{R} \mid \mathbf{O},\mathbf{Z}^{(1)})}_{g(p(\mathbf{R},\mathbf{O},\mathbf{Z}))} = \underbrace{\frac{p(\mathbf{O},\mathbf{Z}^{(1)},\mathbf{R}=1)}{p(\mathbf{R}=1 \mid \mathbf{O},\mathbf{Z}^{(1)})}}_{g(p(\mathbf{R},\mathbf{O},\mathbf{Z}))|_{\mathbf{R}=1}} \times \underbrace{p(\mathbf{R} \mid \mathbf{O},\mathbf{Z}^{(1)})}_{g(p(\mathbf{R},\mathbf{O},\mathbf{Z}))}.$$

请注意，这个泛函与我们之前为因果模型概述的基于逆概率加权的估计量如出一辙，其中缺失机制充当逆权重。

举个简单的例子，图2(d)所示的随机缺失缺失数据有向无环图模型中的完整联合分布可识别，因为缺失机制 $p(R_A,\ R_Y|\mathbf{C}\ A^{(1)},\ Y_{(1)}) = p(R_A,\ R_Y|\mathbf{C})$ 是观测数据的函数。

完全随机缺失、随机缺失与非随机缺失机制的层级之所以与识别相关，仅仅是因为如果缺失性为非随机缺失，完整联合分布可能无法识别。因此，在第3节给出示例和图时，我们通常会指出其中起作用的是哪种机制。

3. 纠缠缺失的图形表示。在讨论了缺失数据和干扰的现有模型后，我们现在提出两种现象共同发生的三类纠缠。通过为这些现象命名并加以描述，我们希望避免在存在干扰的情境中通常采用的做法，即直接忽略缺失数据。

以下定义适用于概率分布 $p(\mathbf{R}, \mathbf{O}, \mathbf{Z}^{(\mathbf{1})}, \mathbf{Z})$，其中所有隐变量均已边缘化。

1. 目标分布依赖（$\mathcal{E}_1$）：区块内个体 i 的反事实变量 $\mathbf{Z}_i^{(1)}$ 和 $\mathbf{Z}_j^{(1)}$ 相互依赖。这意味着目标分布 $p(\mathbf{O}, \mathbf{Z}^{(1)})$ 无法分解为单位特定边缘分布。这种情况出现在以下问题中：如果不存在缺失数据，底层分布会呈现数据依赖或干扰。

2. 缺失过程依赖（$\mathcal{E}_2$）：个体 $\mathbf{R}_i$ 的缺失指示变量 i 依赖于对应个体 j 的变量，因此缺失过程 $p(\mathbf{R} \mid \mathbf{O}, \mathbf{Z}^{(\mathbf{1})})$ 无法分解为单位特定因子。因果推断应用中也研究过类似情况 [80]。3. 缺失干扰（$\mathcal{E}_3$）：个体 i 的观测代理变量 $\mathbf{Z}_i$ 依赖于个体 j.[3] 的缺失指示变量 $\mathbf{R}j$。在 $\mathcal{E}_3$ 下，与变量Z_i 对应的反事实变量不仅必须按被设为 1 的 R_{Z_i} 索引，还必须按其他个体（的变量）的缺失状态索引。例如，反事实变量 $Z^{(R_{Z_i}=1, R_{Z_j}=0)}$ $_i$ 表示个体 i 的变量 Z 在被观测到（$R_{Z_i}=1$），且个体 j 的变量 Z 缺失时的情形。

这三种纠缠可以以任意组合出现，至少出现一种纠缠的七种可能情形共同涵盖了数据依赖与缺失性同时存在的所有情形。为详细描述这些情形，我们使用记号 $\mathcal{E}_i$ 表示第 i 种纠缠存在，并用 $\bar{\mathcal{E}}_i$ 来表示其不存在。例如，我们将用 $\mathcal{E}_1\bar{\mathcal{E}}_2\bar{\mathcal{E}}_3$ 表示这样一种情形：存在目标分布依赖，但不存在缺失过程依赖或缺失干扰。

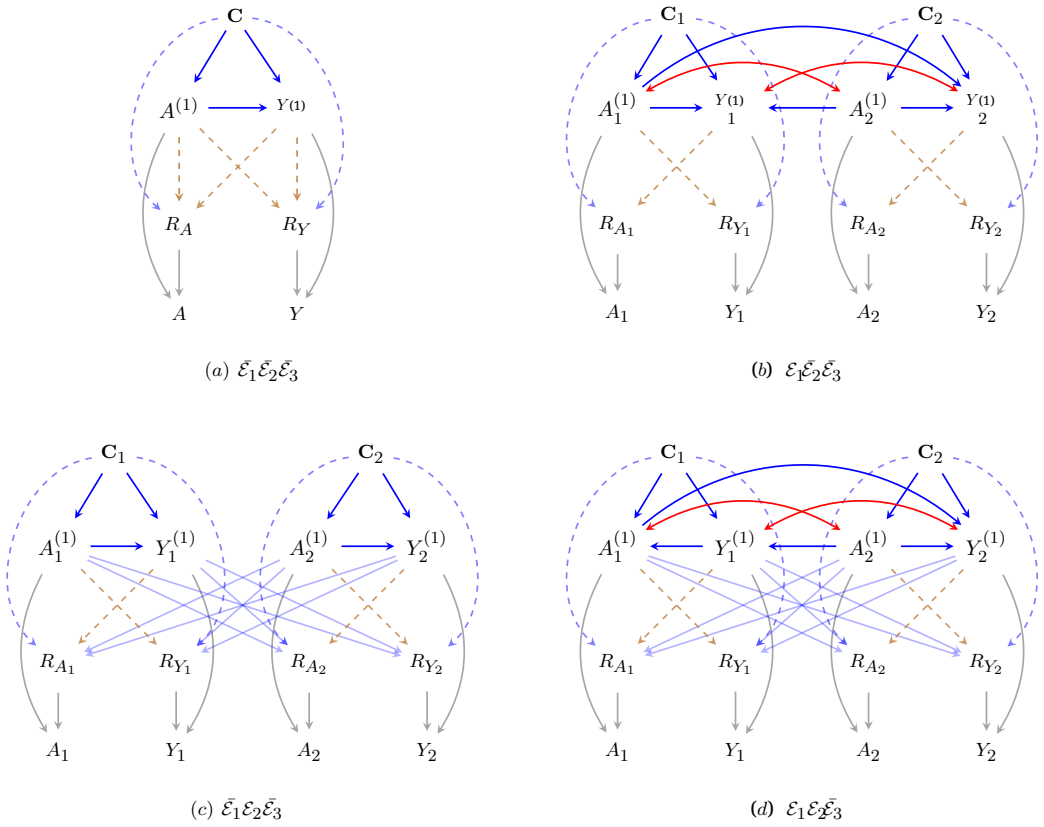
我们首先在第 *3.1* 节介绍不含缺失干扰的纠缠，因为事实证明，这些形式的纠缠与现有的独立同分布缺失数据模型具有相同的图形表示。在第 3.2 节，我们介绍存在缺失干扰的情形；这些情形需要建立新的记号和图形表示。

3.1. 不含缺失干扰的纠缠。 考虑使用社区医院数据库研究药物A 对疾病状态 Y 的效应。假设医院记录不完整，缺失指示变量 R_{A_i}， R_{Y_i} 表示患者 i的处理值和疾病状态值是否有记录，反事实 $A_i^{(1)}$ $Y_i^{(1)}$，表示这些变量的真实（但可能未记录）取值。基线协变量记为 $\mathbf{C}_i$，表示个体 i的协变量，始终可观测，其中包括年龄和饮酒量。在本例中，$\mathbf{Z} \equiv \{A, Y\}$和 $\mathbf{O} \equiv \{\mathbf{C}\}$。为便于说明，我们将使用二元组。

为了对比独立同分布缺失数据模型与存在纠缠的情形，我们首先从一个不涉及任何纠缠的场景$(\bar{\mathcal{E}}_1\bar{\mathcal{E}}_2\bar{\mathcal{E}}_3)$开始，并在此基础上逐步构建示例，以说明纠缠可能如何产生。

情景 1.1（$\bar{\mathcal{E}}_1\bar{\mathcal{E}}_2\bar{\mathcal{E}}_3$）。当所研究的疾病不具传染性，并且与无应答、数据收集缺陷等有关的缺失过程不表现出患者间依赖时，我们或许可以假设，一位患者的记录不会影响任何其他患者的记录，即数据是独立同分布的。

[3]这可以理解为经典干扰的一种扩展：一个个体的处理会影响另一个个体的结局，只不过这里处理被缺失指示变量取代。因此称为缺失干扰。



(a) $\bar{\mathcal{E}}_1\bar{\mathcal{E}}_2\bar{\mathcal{E}}_3$ (b) $\mathcal{E}_1\bar{\mathcal{E}}_2\bar{\mathcal{E}}_3$

(c) $\bar{\mathcal{E}}_1\mathcal{E}_2\bar{\mathcal{E}}_3$ (d) $\mathcal{E}_1\mathcal{E}_2\bar{\mathcal{E}}_3$

图3：四种情景展示了在不存在缺失干扰的二元部分干扰设置中，目标分布依赖和缺失过程依赖可能出现的所有方式。

图 3(a) 描绘了这种情况。若在图 3(a) 中纳入不同的边集，则可得到完全随机缺失，其中 R_A，$R_Y \perp\!\!\!\perp A^{(1)}$，$Y_{(1)}$，$\mathbf{C}$（仅纳入实线时）；随机缺失，其中 R_A，$R_Y \perp\!\!\!\perp A^{(1)}$，$Y_{(1)} | \mathbf{C}$（除实线外还纳入蓝色虚线边时）[4]；或非随机缺失（至少纳入一条棕色虚线边时）。边 $A \to R_A$ 和$Y \to R_Y$称为自我删失或自我遮蔽边 [12, 81, 82]。

情景1.2（$\bar{\mathcal{E}}_1\mathcal{E}_2\mathcal{E}_3$）。假设结局 Y 是一种传染病，例如 COVID-19。由于患者可能相互传染，居住在同一家庭或地理区域的患者数据很可能在结局上存在依赖。此外，感染激增可能导致治疗短缺，进而因分配干扰 [64]导致治疗之间存在依赖；在COVID-19大流行早期的多个时期，都曾出现地域集中的治疗短缺。最后，对个体 i 的传染性疾病结局进行成功预防或治疗，可能通过阻止其向其他个体传播，影响其他个体 j 的结局；也就是说，阻止疾病从 i 传播给 j。尽管存在这些复杂情况，各个体之间的缺失机制仍然相互独立。

图 3（b）展示了这一情形。双向箭头表示 $A_1^{(1)}$ 与 $A_2^{(1)}$ 之间，以及 $Y_1^{(1)}$ 与 $Y_2^{(1)}$ 之间的依赖（由共同的隐藏原因导致）。为简化图示，我们省略了 $\mathbf{C}_i$ 与 $\{A_j, Y_j\}$之间的连接。图 3（b）

[4]请注意，随机缺失版本与图 2（d）相同，因为情景1.1的所有内容均适用于独立同分布设置。

仅由实线边组成的子图对应完全随机缺失模型，因为 $\mathbf{R} \perp\!\!\!\perp \mathbf{Y}^{(1)}, \mathbf{A}^{(1)}, \mathbf{C}$ 根据 m-分离准则 [58]。同样，图3(b)中由除棕色虚线边以外的所有边组成的子图对应随机缺失模型，因为 $\mathbf{R} \perp\!\!\!\perp \mathbf{Y}^{(1)}, \mathbf{A}^{(1)} \mid \mathbf{C}$。最后，在随机缺失模型中加入棕色虚线边，就得到非随机缺失模型。

情景 1.3 ($\mathcal{E}_1\mathcal{E}_2\mathcal{E}_3$)。 假设我们再次讨论一种非传染性疾病，并且某个个体的治疗分配不受其他个体任何特征的影响。然而，一个个体的缺失过程确实依赖于其他个体。例如，数据包含同一个家庭的成员，他们可能共同决定更换医疗服务提供者，从而导致该家庭所有成员的治疗与结局数据缺失。

图 3(c) 描绘了这一情形。反事实 $A_1^{(1)}, Y_{(1)1}$与缺失指示变量 R_{A_2}, R_{Y_2}之间的浅蓝色边，以及 $Y_{(1)2}, Y_{(1)2}$与 R_{A_1}, R_{Y_1}之间的浅蓝色边，表示缺失过程之间存在依赖。为简洁起见，省略了从 $\mathbf{C}_2$ 指向个体 1变量的边，以及从 $\mathbf{C}_1$ 指向个体 2 变量的边。该情景的完全随机缺失版本仅简化为情景 $\mathcal{E}_1\mathcal{E}_2\mathcal{E}_3$，代表独立同分布数据。此情景中的一个随机缺失机制示例，对应于图 3(c) 中没有浅蓝色实线边，而是存在边$\mathbf{C}_1 \to R_{A_2}$, $\mathbf{C}_1 \to R_{Y_2}$, $\mathbf{C}_2 \to R_{A_1}$和 $\mathbf{C}_2 \to R_{Y_1}$（图中未示出）。缺失过程之间的依赖也可以表示为个体 i 的变量与个体 j的缺失指示变量之间的一条双向边，该边源于相应变量与缺失指示变量的潜在共同原因。

情景 1.4（$\mathcal{E}_1\mathcal{E}_2\mathcal{E}_3$）。此情景简单地结合了情景 1.2 和 1.3 中的复杂情况，因此目标分布依赖和缺失过程依赖同时存在。边的解释与情景 1.2 和 1.3 中完全相同。具体而言，反事实与缺失指示变量之间的浅蓝色实线边使缺失过程依赖属于非随机缺失。相反，如果这种依赖存在于观测协变量上，则该机制属于随机缺失。棕色和蓝色虚线边均作用于一个个体内部，分别诱导属于非随机缺失或随机缺失的依赖，且不涉及其他个体。

最后，我们注意到，图 3 中的图可以直接扩展到这样的网络：其中区块不再只是二元组，而是具有任意大小 m，以说明本文讨论的不同类型的纠缠。与截至目前所述设置相对应的识别结果将在第4.1节讨论。

3.2. 缺失性与干扰的纠缠。 在某些缺失性与数据依赖同时出现的情形中，底层变量的观测实现值会受到多个个体的缺失指示变量影响，从而形成缺失数据领域中因果推断干扰问题的对应问题。

考虑以下示例：研究人员希望根据参与者的基线健康指标，评估一项健康与 wellness 项目的结局，该项目包括饮食与锻炼方案。参与者通过自愿登记参加研究。Bob 和 Anne 是邻居，他们看到了招募广告。设想两种不同的情形：（1）Bob 和 Anne 都对该项目充满信心，并一同入组。由于一起参与，他们能够轻松规划膳食和身体活动，并互相鼓励。因此，Anne 取得了成功的结局：新陈代谢得到改善，身体质量指数也有所提高。另一种情形是，（2）Bob 入组并收集了他的基线变量，但实际上并未参与。Anne 完成了项目，但发现很难保持动力并有效遵循方案，她的结果也没有那么好。由于 Bob 没有参与，因此没有测量他的最终结局（身体质量指数）。在这个例子中，Anne 有两种可能的

假设性项目结果，分别对应上述两种情形：一种是 Bob 参与，另一种是他没有参与。换句话说，Anne 的观测结果取决于 Bob 在研究中的结果的缺失状态。对 Bob 也可以作出类似的论述。

形式上，对于 Anne（索引为 1 的个体），结局身体质量指数有两个可能值：Y（$r_{Y_1}=1,r_{Y_2}=0$）1 表示 Bob 不参与，因此其身体质量指数未被记录（$r_{Y_2}=0$），$Y^{(r_{Y_1}=1,r_{Y_2}=1)}$）1 表示 Bob 参与，因此其身体质量指数被记录（$r_{Y_2}=1$）。这里我们做了两项简化。第一，我们假设对于任何个体，唯一的缺失变量是结局，因此将 r_{Y_1} 记为 r_1，将 r_{Y_2} 记为 r_2。这一假设仅为简化本例而作，在更一般的情形中将予以放宽（见第 3.2.1 节）。第二，我们从上标中省略 r_1，并将反事实记为 $Y^{(1,r_2=0)}$）1 和 $Y^{(1,r_2=1)}$）1。后一项简化不会造成歧义，因为只有在 $r_1=1$ 时才定义 Anne 的结局。Anne 的缺失状态 R_1 和可观测的身体质量指数 Y_1 均记录在研究数据中。这个观测结果不仅取决于她的反事实 $Y^{(1,r_2=0)}$）1、$Y^{(1,r_2=1)}$）1 及其缺失状态 R_1，还取决于 Bob 的结局缺失状态 R_2。可以将 R_2 视为一个开关，用来选择在观测中实现的是哪个反事实：$Y^{(1,r_2=0)}$）1 或 $Y^{(1,r_2=1)}$）1。在这个特定例子中，基于对称性，我们假设 Bob 也满足同样的关系。因此，邻居 Bob 和 Anne 的观测值 Y_1、Y_2 与相应反事实之间存在以下关系：

$$Y_1 \leftarrow \begin{cases} (1-r_2)Y_1^{(1,r_2=0)} + r_2 Y_1^{(1,r_2=1)} & \text{if } r_1 = 1 \\ ? & \text{if } r_1 = 0 \end{cases}$$

$$Y_2 \leftarrow \begin{cases} (1-r_1)Y_2^{(1,r_1=0)} + r_1 Y_2^{(1,r_1=1)} & \text{if } r_2 = 1 \\ ? & \text{if } r_2 = 0 \end{cases}$$

由于此情形涉及多个关于 Y_1的反事实版本，因此无法用迄今描述过的缺失数据模型来刻画，也无法用图 3 中的图来描述；这些图对应的是每个变量都只有一个反事实版本的完整数据分布。

对于每个个体 i，我们将其缺失指示变量所指向的、对应于个体 i 的反事实的个体集合称为网络中 i 的影响者，记作 aff(i)。这一术语的使用受神经科学 [83] 中用法的启发。可以想见，这种关系不一定是对称的；一个个体可以是另一个个体的影响者，但反过来不一定成立。考虑到这些因素，我们现在描述缺失干扰可能出现的四种情景。

情景 2.1（$\mathcal{E}_1\mathcal{E}_2\mathcal{E}_3$）。首先，考虑这样一种情形：既不存在目标分布依赖，也不存在缺失过程依赖，只有缺失干扰。假设根据参与者测得的基线协变量来推荐饮食和锻炼方案。作为一种近似，我们假设一个人的身体质量指数只受其自身行为影响，因此目标分布不会彼此直接影响。如前所述，如果 Anne 的朋友 Bob 锻炼（$R_j \to Y_i$），Anne 更有可能锻炼，从而影响她的身体质量指数。我们假设，对于任意个体 $Y^{(1,r_j=0)}{}_i$ 和 $Y^{(1,r_j=1)}{}_i$ 之间存在关联，这种关联是由潜在共同原因 i 导致的H_i，从而在潜在投影图中，反事实之间形成一条双向(↔）边。见图 4(a)。棕色虚线边表示自我删失机制。

情景 2.2 ($\mathcal{E}_1\mathcal{E}_2\mathcal{E}_3$**)**。现在，假设 Bob 和 Anne 是兄弟姐妹。可以设想，饮食和锻炼对个体的效应由他们共同拥有的遗传因素介导，从而产生目标分布依赖。此外，假设负责这项

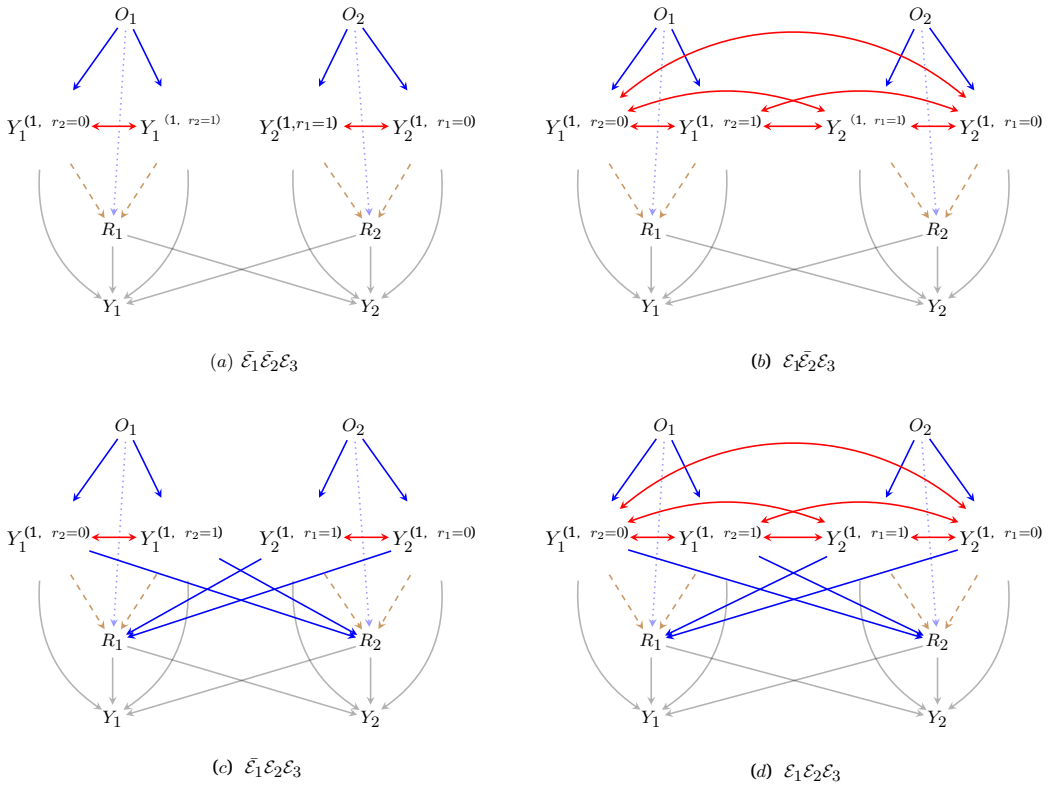


图 4：在存在缺失干扰的二元部分干扰设置中，目标分布依赖和缺失过程依赖可能出现的所有情形的四种示意。我们假设只有结局 Y_1，Y_2 缺失，并用 R_i 作为 R_{Y_i} 的简写。

研究的人员会从登记的参与者名单中随机决定谁可以参加（因此，某人是否被纳入研究与其兄弟姐妹无关）。这会形成图 4(b) 所示的结构。对规定方案的依从性仍取决于某人的兄弟姐妹是否被纳入研究，从而使每个个体都有两个潜在的反事实 BMI。这里同样假设，反事实之间的依赖是由共同原因 H 导致的，目标分布依赖则由双向边表示。

情形 2.3（$\bar{\mathcal{E}}_1\mathcal{E}_2\mathcal{E}_3$）。假设 Bob 和 Anne 没有遗传关系。此外，假设得知 Bob 入组（或未入组）时的潜在 BMI 会使 Anne 更有可能自行入组，并且研究人员纳入所有登记者，而不随机抽取部分登记者，因此产生了缺失过程依赖。由此得到的图示如图 4(c) 所示。蓝色的边 $Y_{(1,\ r_1=1)_2} \to R_1$ 引入了一种更一般的自我删失类型，我们称之为 影响者删失。这种情形对应于：某个个体的反事实变量（按其影响者的缺失状态索引）对该个体自身的缺失状态进行删失。

情景 2.4（$\mathcal{E}_1\mathcal{E}_2\mathcal{E}_3$）。最后，假设这样一种情形：Bob 和 Anne 具有共同的遗传因素，并且根据他们对另一人在研究过程中可能达到何种健康水平的了解，他们也更有可能入组。这一设定将三种可能的纠缠类型结合在一起。该情景通过图 4(d) 中的图示呈现。

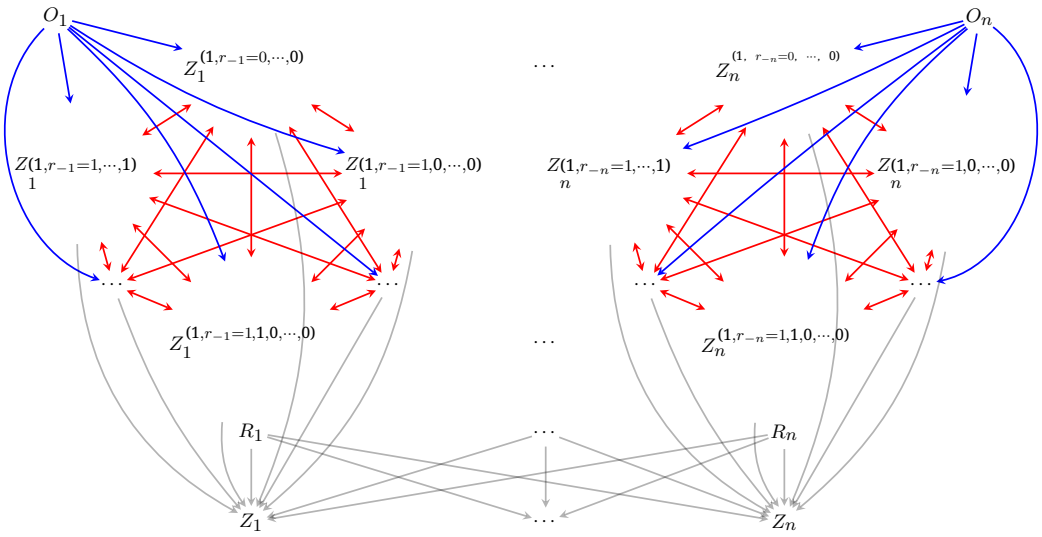


图5：一个简单的一般网络示例，其中有 n 个体受到缺失干扰。该图表示一个完全随机缺失模型，$\mathbf{R} \perp\!\!\!\perp \mathbf{O}$，$\mathbf{Z}^{(1,\mathbf{r}_{\text{aff}(i)})}{}_i$。

3.2.1. 超越二元组。在许多现实情境中，相互作用并不局限于二元组，因此我们描述存在缺失干扰的网络的反事实（和图）。为便于说明，我们作如下假设：给定个体 i 和 j，若 $R_{Z_j} \to Z_i$ 存在，则对所有 $Z'_j \in \mathbf{Z}j$，$R_{Z'_j} \to Z_i$ 也存在。我们将这一假设记为 $j \in \text{aff}(i)$。这样，我们就可以用 $\mathbf{R}j$ 为个体 i 的反事实编索引；$\mathbf{Z}j$ 用于确定其是否缺失，而不是单个变量 $Z_j \in \mathbf{Z}j$。我们进一步假设，在同一个体内，对于任意不同的变量 Z, Z'，边 $R_{Z_i} \to Z'_i$ 均不存在。这样便可以在个体 i 的反事实索引中，将 $R_{Z_i} = 1$ 替换为 1。若要了解取消这些限制的影响，请参阅本节末尾的注释。

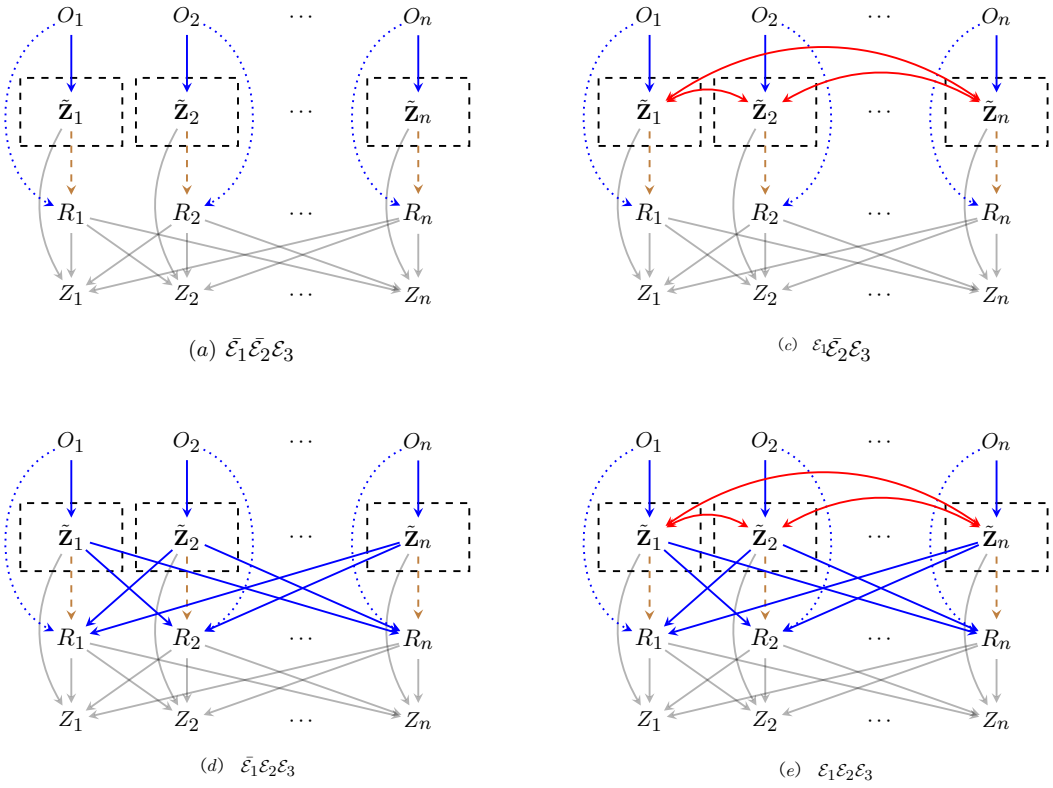
˜我们假设一个由 n 个个体组成的网络。完整联合分布定义在$\bigcup_{i=1}^n \mathbf{O}_i \cup \mathbf{Z}_i \cup \mathbf{R}_{\mathbf{Z}_i} \cup \mathbf{H}_i$ 中的变量上，其中 $\tilde{\mathbf{Z}}_i \equiv \{\mathbf{Z}^{(1,\mathbf{r}_{\text{aff}(i)})}{}_i : \mathbf{r}_{\text{aff}(i)}$ R的一个取值$_{\text{aff}(i)}\}$ 表示个体 i的变量$\mathbf{Z}_i$ 的反事实版本；aff(i)表示作为个体 i影响者的个体集合；而$\mathbf{R}_{\text{aff}(i)} \equiv \{\mathbf{R}^{\mathbf{Z}_j} : \ j \in \text{aff}(i)\}$，其中$\mathbf{R}^{\mathbf{Z}_j}$指个体 j的所有变量 $Z_j \in \mathbf{Z}j$ 对应的全部缺失指示变量 R_{Z_j} 的集合。换言之，对于每个取值$\mathbf{r}^{\text{aff}(i)}$，变量$\mathbf{Z}^{(1,\mathbf{r}_{\text{aff}(i)})}{}_i \in \tilde{\mathbf{Z}}_i$表示“假如变量$\mathbf{Z}_i$曾被观测到，并且网络中 i 的影响者的可观测状态被设为$\mathbf{r}^{\text{aff}(i)}$，那么这些变量所取的值”。

作为（2）的推广，每个观测代理变量 $Z_i \in \mathbf{Z}_i$ 定义如下：

$$Z_i \leftarrow \begin{cases} Z_i^{(1,\mathbf{R}_{\text{aff}(i)})} & \text{if } R_{Z_i} = 1 \\ ? & \text{if } R_{Z_i} = 0 \end{cases}, \tag{6}$$

其中，上标中的 $\mathbf{R}^{\text{aff}(i)}$ 表示将 $\mathbf{R}^{\text{aff}(i)}$ 设为其自然发生值的干预。与因果干扰一样，通常情况下，对应于同一个变量 Z_i 的所有反事实彼此依赖。接下来，我们讨论这些网络的图。

完整数据分布可以用一个有向无环图表示，该图为每个$Z^{(1,\mathbf{r}_{\text{aff}(i)})}{}_i$ 以及$\bigcup_{i=1}^n \mathbf{O}_i \cup \mathbf{Z}_i \cup \mathbf{R}_{\mathbf{Z}_i} \cup \mathbf{H}_i$中的每个元素都包含一个顶点。我们还假设，同一个体 i 的变量 Z_i 的所有反事实共享一个共同的父节点 $H_{Z_i} \in \mathbf{H}_i$，因为它们



(a) $\bar{\mathcal{E}}_1\bar{\mathcal{E}}_2\mathcal{E}_3$ (c) $\mathcal{E}_1\bar{\mathcal{E}}_2\mathcal{E}_3$

(d) $\bar{\mathcal{E}}_1\mathcal{E}_2\mathcal{E}_3$ (e) $\mathcal{E}_1\mathcal{E}_2\mathcal{E}_3$

图6：一个具有 n 个体且存在缺失性干扰的一般网络。(a)–(d)列举了目标分布依赖和缺失过程依赖可能出现的所有方式，是图中二元组的推广。4。

彼此相关。该有向无环图遵循两个限制，这些限制推广了第 2.3 节讨论的限制：(i) $\mathbf{R}$ 中的变量不能指向$\bigcup_{i=1}^{n}\mathbf{O}_i \cup \{\mathbf{Z}^{(1,\mathbf{r}_{\text{aff}(i)})}{}_i$ }中的变量；(ii) Z_i 位于$\mathbf{Z}$ 中的每个元素都以其所有对应反事实、的影响者以及 R_{Z_i} 为父节点。限制 (ii) 是式 (6) 中一致性性质的结果。

将此有向无环图中的完全隐藏变量$\bigcup_{i=1}^{n}\mathbf{H}_i$ 投影出去得到的潜在投影无向混合图，将是我们首选的图形表示方式，与图3和图4中的做法相同。图5展示了一般网络中此类祖先有向混合图的一个简单示例，其中每个个体 i只有一个缺失变量 Z_i。反事实之间的依赖$\mathbf{Z}^{(1,\mathbf{r}_{\text{aff}(i)})}{}_i$ 由双向边表示，这些边是通过将 H_i 投影出去得到的。在此示例中，每个个体都是其他每个个体的影响者。为使表示简洁，我们只展示了其中两个个体的反事实。

当图中顶点 $Z^{(1,\mathbf{r}_{\text{aff}(i)})}{}_i$ 和 $Z^{(1,\mathbf{r}'\text{aff}(i))}{}_i$ 对于所有 $\mathbf{r}$、$\mathbf{r}'$ 的入边和出边都相同时，我们可以引入对应于集合 $\tilde{}\mathbf{Z}_i$ 的顶点 $\tilde{}\mathbf{Z}_i$，以替代其所有元素 $Z^{(1,\mathbf{r}_{\text{aff}(i)})}{}_i$，并适用于所有可能的、对应于 $\mathbf{r}_{\text{aff}(i)}$的$\tilde{}$索引方式，从而使图形表示更加紧凑。我们将这类图记为 $\mathcal{G}$，如图 6 所示。特别是，图 6(a) 中由实线边组成的图是图 5 中图的压缩表示；任何顶点 $\mathbf{Z}_i$（以黑色虚线框表示）都紧凑地表示由双向边相连的完整顶点集合 $\{Z_i^{(1,\mathbf{r}_{\text{aff}(i)}=0,\cdots,0)}, Z^{(1,\mathbf{r}_{\text{aff}(i)}=1,\cdots,0)}{}_i, \cdots, Z^{(1,\mathbf{r}_{\text{aff}(i)}=1,\cdots,1)}{}_i$ }。在 $\mathcal{G}$ 中，对于任意节点 V，我们将边 $V \to \mathbf{Z}_i$解释为表示 V

影响 $\tilde{\mathbf{Z}}_i$ 中的所有反事实 $\mathbf{Z}^{(1,\mathbf{r}_{\text{aff}(i)})}{}_i$。类似地，$V \leftarrow \tilde{\mathbf{Z}}_i$ 表示 $\tilde{\mathbf{Z}}_i$ 中的所有反事实 $\mathbf{Z}^{(1,\mathbf{r}_{\text{aff}(i)})}{}_i$ 都会影响 V。

此外，将 $\mathcal{G}$ 视为图6(a) 的广义版本很有帮助。图4(a) 中不存在目标分布依赖或缺失过程依赖，只有缺失干扰。与之前一样，该图可以表示完全随机缺失（仅有实线边）、随机缺失（有蓝色虚线边）和非随机缺失（有棕色虚线边）过程。图6(b)-(d) 是图4(b)-(d) 中各图的广义版本。图 $\mathcal{G}$ 适合用于说明，因为它们简洁明了；但正如我们在第4节所示，在反事实不共用相同边的模型中，识别更加直接。因此，在本文余下部分，我们将使用图 $\mathcal{G}$，例如图5 中的图（而非 $\mathcal{G}$ 图6中的图）。

我们考察过这样的例子：某个个体中所有可能缺失的变量要么全部观测到，要么全部缺失，这意味着每个个体只需要一个缺失指示变量。一般而言，情况未必如此。这会带来两种不同的复杂情形：(i) 对于某些个体 R_{Z_j}，可能存在从 Z_i' 指向变量 i的边，j 以及变量 Z、$Z' \in \mathbf{Z}$；(ii) 也可能存在从某个个体的缺失指示变量指向该个体内另一个变量的边（例如 $R_{Z_i} \to Z_i'$）。若出现情形 (i)，我们就不能像之前那样仅依据其他个体的缺失指示变量为反事实编索引，而必须依据这些个体的特定变量。为与先前的定义保持一致，如果某个个体的任一缺失指示变量为另一个个体的反事实编索引，我们就称前者为后者的影响者。若出现情形 (ii)，缺失干扰发生在个体内部，而非个体之间；在不存在其他相互作用时，数据仍然是独立同分布的。关于这一问题的更多讨论，请参见第5 节。值得指出的是，在这两种情况下，图结构和数学框架仍与本节迄今讨论的相同。变化的是个体与其记录变量之间关系的解释。

4. 纠缠缺失模型中的识别。 在建立了纠缠的符号和图模型框架后，下面我们讨论这些模型中的识别问题。

4.1. 无缺失干扰时的识别。 我们在第 3.1 节中指出，在不存在缺失干扰的情况下，（图）模型与现有的缺失数据独立同分布模型基本相同，唯一的区别在于我们如何看待最小研究单位——它是单个个体，还是区块内相互作用的个体集合。

在缺失数据文献中，通常将形如 $\beta = \mathbb{E}[h(\mathbf{Z}^{(1)}, \mathbf{O})]$, 的期望视为感兴趣的参数，其中 h 是 $\mathbf{Z}^{(1)}$、$\mathbf{O}$ 的某个已知函数。然而，本文将按照因果图模型识别文献的惯例，讨论分布的识别问题。具体而言，我们关注完整联合分布 $p(\mathbf{Z}^{(1)}、\mathbf{O}、\mathbf{R})$ 的非参数识别，因为如果完整联合分布能够由观测数据实现非参数识别，那么它的任何泛函也同样能够被识别。

一个关于在具有完全观测变量的独立同分布缺失数据有向无环图模型中进行完整联合分布识别的可靠且完备算法已在 [12] 中提出。[5] 作者还在同一项工作中对具有隐藏变量的缺失数据模型进行了完整联合分布识别。对于 DAG 和祖先有向混合图这两类图，其识别准则依赖于马尔可夫毯的概念 [54, 62]：在一个祖先有向混合图中 $\mathcal{G}$，顶点 V 的马尔可夫毯，记作 $\text{mb}_{\mathcal{G}}(V)$，由所有与 V 共享一条边或通过碰撞路径与 V 相连的变量组成；一条碰撞路径

[5] 可靠且完备的算法对应于必要且充分的识别假设。

是指路径上的所有顶点均为以下形式的碰撞点的路径： $\rightarrow o \leftarrow$、 $\leftrightarrow o \leftarrow$、 $\leftrightarrow o \leftrightarrow$。在有向无环图（一种没有双向边的祖先有向混合图）中，马尔可夫毯缩减为与 V 有共同边的变量集合，以及与 V 共享子节点的变量。我们简要介绍其结果，见下文定理 1 和定理 2，因为这些结果可直接应用于不存在缺失干扰的纠缠缺失情形。

定理 1. [有向无环图中的完整分布识别 [12]]

在由有向无环图 $\mathcal{G}$ 表示的缺失数据模型中，若且唯若 $Z^{(1)} \notin \mathrm{mb}_{\mathcal{G}}(R_Z)\ \forall Z^{(1)} \in \mathbf{Z}^{(1)}$，完整联合分布 $p(\mathbf{R}, \mathbf{Z}^{(1)}, \mathbf{O})$ 可识别。因此，要使完整联合分布可识别，就不能存在形如 $Z^{(1)} \rightarrow R_Z$ 的边（即不存在自我删失），也不能存在形如$Z^{(1)} \rightarrow R_{Z'} \leftarrow Z$ 的结构（即不存在共谋点）。识别泛函由式5给出，其中缺失机制 $p(\mathbf{R} \mid \mathbf{Z}^{(1)}, \mathbf{O})$ 由优势比参数化[84]给出：

$$(7)\quad \frac{1}{\sigma(\mathbf{Z}^{(1)}, \mathbf{O})} \times \prod_{k=1}^{K} p(R_k \mid \mathbf{R}_{-k} = 1, \mathbf{Z}^{(1)}, \mathbf{O}) \times \prod_{k=2}^{K} \mathrm{OR}(R_k, R_{\prec k} \mid R_{\succ k} = 1, \mathbf{Z}^{(1)}, \mathbf{O}),$$

其中$\mathbf{R}_{-k} = \mathbf{R} \setminus R_k, R_{\prec k} = \{R_1, \cdots, R_{k-1}\}, R_{\succ k} = \{R_{k+1}, \cdots, R_K\}$,

$$\mathrm{OR}(R_k, R_{\prec k} \mid R_{\succ k} = 1, \mathbf{Z}^{(1)}, \mathbf{O}) = \frac{p(R_k \mid R_{\succ k} = 1, R_{\prec k}, \mathbf{Z}^{(1)}, \mathbf{O})}{p(R_k = 1 \mid R_{\succ k} = 1, R_{\prec k}, \mathbf{Z}^{(1)}, \mathbf{O})} \times \frac{p(R_k = 1 \mid \mathbf{R}_{-k} = 1, \mathbf{Z}^{(1)}, \mathbf{O})}{p(R_k \mid \mathbf{R}_{-k} = 1, \mathbf{Z}^{(1)}, \mathbf{O})},$$

且 σ（$\mathbf{Z}^{(1)}$， $\mathbf{O}$）$= \sum_r \{\prod_{k=1}^{K} p$（$r_k|\mathbf{R}_{-k} = 1$， $\mathbf{Z}^{(1)}$， $\mathbf{O}$）$\times \prod_{k=2}^{K}$优势比（r_k, $r_{\prec k}|R_{\succ k} = 1$， $\mathbf{Z}^{(1)}$， $\mathbf{O}$）$\}$是归一化函数。

考虑图2（d）中的独立同分布图。 $p(R_A, R_Y \mid \mathbf{C}\ A^{(1)}, Y(1))$ 的优势比参数化与其常规有向无环图参数化（见第2.3节）相同，因为 $p(R_A \mid R_Y = 1, \mathbf{C}\ A^{(1)}, Y(1)) = p(R_A|\mathbf{C})$、 $p(R_Y|R_A = 1, \mathbf{C}\ A^{(1)}, Y(1)) = p(R_Y|\mathbf{C})$， 且 $\mathrm{OR}(R_A, R_Y \mid Y(1)\, A^{(1)}, \mathbf{C}) = \sum_{R_Y}, {}_{R_A} p(R_A|\mathbf{C}) \times p(R_Y|\mathbf{C}) = 1$，归一化项为1。自我删失边 $A^{(1)} \rightarrow R_A$ 和 $Y(1) \rightarrow R_Y$阻碍了图3（a）中完整联合分布的识别。

下述定理总结了由含隐变量的缺失数据有向无环图的潜在投影所得到的无环有向混合图中的完整分布识别结果。

定理 2. [无环有向混合图中的完整分布识别 [12]]

在由隐变量有向无环图 $\mathcal{G}(\mathbf{R}, \mathbf{Z}^{(1)}, \mathbf{O}, \mathbf{H})$ 及其潜在投影无向混合图 $\mathcal{G}(\mathbf{R}, \mathbf{Z}^{(1)}, \mathbf{O})$ 表示的缺失数据模型中，当且仅当 $Z^{(1)} \notin \mathrm{mb}_{\mathcal{G}}(R_Z)\ \forall Z^{(1)} \in \mathbf{Z}^{(1)}$ 时，完整联合分布 $p(\mathbf{R}, \mathbf{Z}^{(1)}, \mathbf{O})$ 可识别。因此，要使完整联合分布可识别，任意一对 $(Z^{(1)}, R_Z)$ 都不应直接相连（即不存在自我删失），也不应通过碰撞路径相连（即不存在共谋路径）。此外，缺失机制的识别由优势比参数化给出，如定理 1 所述。

图3中没有任何图3包含共谋路径。然而，举例来说，如果路径$A_1^{(1)} \rightarrow R_{Y_1} \leftarrow R_{A_1}$或路径 $A_1^{(1)} \leftrightarrow A_2^{(1)} \leftrightarrow R_{A_1}$ 存在于其中任何一个图中，那么完整联合分布将无法被识别。

存在纠缠但没有缺失干扰的情形适用于这些（现有的）识别结果，但要实现有效识别，必须仔细考虑纠缠所产生的依赖，以及它们是否会在相应的图中引入自我删失或共谋路径。

4.2. 存在缺失干扰时的识别。 在存在缺失干扰的情形下，所有反事实的联合分布（通常称为目标分布）涉及来自多个世界的反事实变量。例如， $Y^{(1,r_2=0)}$ 1 和 $Y_{(1,r_2=1)}$ 1 在第3.2节 Anne 的情形中就是如此。这意味着，如果没有非常强的假设，就无法识别完整联合分布。[6]相反，我们将考虑对单一世界对象的识别。在给出严格定义之前，我们先用一些简单的例子来说明这一想法。对于 Bob 和 Anne，我们可能感兴趣的问题包括："如果 Bob 退出研究，Anne 的身体质量指数会是多少？"或者"如果他们都继续参与研究，各自的结局或联合结局会是什么？"，而不是"如果 Bob 或 Anne 中的另一人不参与研究，他们两人的联合结局会是什么？"，因为相应事件发生在不同的世界中。

对于给定的网络和缺失指示变量 R =r 的赋值，单世界查询定义为反事实变量的边缘分布 $\{\bigcup_{i=1}^{n} Z^{(1,\mathbf{r}_{\text{aff}(i)})}\, i \quad |r_i = 1$ 若个体 i 包含在此集合中$\}$. 例如，考虑一个包含 3 个个体的网络，其中 $\mathbf{r}^{\text{aff}(1)} = (r_2, r_3)$、$\mathbf{r}^{\text{aff}(2)} = (r_1, r_3)$ 和 $\mathbf{r}^{\text{aff}(3)} = (r_1, r_2)$。给定 $\mathbf{r} = (r_1, r_2, r_3) = (1, 0, 1)$ 的取值，我们将查询 $p(Z^{(1,r_2=0,r_3=1)}\, 1\, , Z^{(1,r_1=1,r_2=0)}\, 3\,)$ 视为有效的单世界查询，而任何包含 $Z^{(1,\mathbf{r}_{\text{aff}(2)})}\, 2$ 的边缘分布都是无效的。这是因为反事实 $Z^{(1,r_2=0,r_3=1)}\, 1$ 和 $Z^{(1,r_1=1,r_2=0)}\, 3$ 中的赋值 $r_2 = 0$ 与反事实 $Z^{(1,\mathbf{r}_{\text{aff}(2)})}\, 2$ 中的赋值 $r_2 = 1$ 不一致。我们用 $h(\mathbf{Z}; \mathbf{r})$ 表示缺失数据网络中的单世界˜ ˜ ˜查询，其中 $\mathbf{Z} \equiv \{\mathbf{Z}_i : i \in \{1, \ldots, n\}\}$ 是所有反事实的集合。

我们的第一个识别结果表明，在存在缺失干扰的情况下，完全随机缺失和随机缺失模型的类比模型能够对任何单世界查询h（$\mathbf{Z}$；$\mathbf{r}$）实现非参数识别。

我们概述的第一个识别结果表明，当存在缺失干扰时，与完全随机缺失和随机缺失模型相对应的模型能够对任意单一˜世界对象 $h(\mathbf{Z};\mathbf{r})$实现非参数识别。

定理 3。在具有缺失干扰的缺失数据ADMG $\mathcal{G}$ 中，由一组反事实 $\mathbf{Z}' \equiv \bigcup \mathbf{i}\{\mathbf{Z}^{(\mathbf{1},\mathbf{r}_{\text{aff(i)}})}\, \mathbf{i}\}$，$i \in \{1, \cdots, n\}$ 构成的有效单世界对象 $h(\tilde{\mathbf{Z}}; \mathbf{r})$ 在满足以下两个条件之一时可识别：(1) $\mathbf{R}' \perp\!\!\!\perp \mathbf{O}, \mathbf{Z}$（完全随机缺失），或 (2)˜$\mathbf{R}' \perp\!\!\!\perp \mathbf{Z}|\mathbf{O}$（随机缺失），其中 $\mathbf{R}'$ 指代所有缺失指示变量的集合 R，这些变量在 $h(\mathbf{Z}; \mathbf{r})$ 中为反事实编制索引。对象 $h(\mathbf{Z}; \mathbf{r})$ 是 $p(\mathbf{Z}', \mathbf{R}, \mathbf{O})$ 的函数，识别泛函为：

$$p(\mathbf{Z}', \mathbf{R}, \mathbf{O}) = p(\mathbf{Z}', \mathbf{O}) \times p(\mathbf{R}|\mathbf{O}, \mathbf{Z}') = \frac{p(\mathbf{Z}', \mathbf{R} = \mathbf{r}, \mathbf{O})}{p(\mathbf{R} = \mathbf{r} \mid \mathbf{O})} \times p(\mathbf{R} \mid \mathbf{O}) \tag{8}$$

其中，倾向评分是通过对祖先有向混合图（或有向无环图）的因子分解应用简单的 m-分离或（d-分离）规则得到的。

[6]参见 [16, 56]，了解秩保持假设如何关联多个世界中的反事实。

证明：证明见附录。 □

在存在MNAR缺失机制时，单世界查询未必总是可识别的。我们转而考虑一种特殊的单世界查询，称为完全可观测性定律，$p(\tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})}, \mathbf{R})$，其中 $\tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})} \equiv \bigcup_{i=1}^{n} Z_i^{(1,\mathbf{r}_{\text{aff}(i)}=\mathbf{1})}$ 是所有将缺失状态设为完全可观测性的反事实集合（即，缺失模式为$\mathbf{R} = \mathbf{1}$）。完全可观测性定律可视为与独立同分布设置下的完整联合分布十分相似。它对应于 Anne 和 Bob 示例中的分布 $p(Y_{(1,\ r_2=1)\,1},\ Y_{(1,\ r_1=1)\,2},\ R_1,\ R_2)$。事实上，在某些假设下，完全可观测性定律是可识别的。在介绍如何在具有纠缠缺失性的非随机缺失模型中识别完全可观测性定律这一结果之前，我们先概述主要假设和几个定义。

我们首先假设，能够作为任意 $R_i \in \mathbf{R}$ 父节点的反事实类型只能具有 $Z^{(1,\mathbf{r}_{\text{aff}(j)}=\mathbf{1})}{}_j$ 的形式。我们将这一假设形式化如下。

假设1。在具有缺失干扰的缺失数据ADMG中，$\mathcal{G}\,\text{ch}_{\mathcal{G}}(\tilde{\mathbf{Z}}^{(\mathbf{r}\neq\mathbf{1})}) \cap \mathbf{R} = \emptyset$。这里，$\tilde{\mathbf{Z}}^{(\mathbf{r}\neq\mathbf{1})}$ 表示所有对应于至少包含一个零的模式的反事实集合，即 $\tilde{\mathbf{Z}}^{(\mathbf{r}\neq\mathbf{1})} = \tilde{\mathbf{Z}} \setminus \tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})}$。此外，定义 $\text{ch}_{\mathcal{G}}(V)$ 以析取方式适用于集合 $\mathbf{V}$，即 $\text{ch}_{\mathcal{G}}(\mathbf{V}) = \bigcup_{V\in\mathbf{V}} \text{ch}_{\mathcal{G}}(V)$。

接下来，我们定义以下新的图结构，以用于我们的结果。

1. *e*-自我删失 [7] 或 影响者删失：对于给定的 $R_k \in \mathbf{R}^{\text{aff}(i)}$，若 $Z^{(1,\mathbf{r}_{\text{aff}(i)}=\mathbf{1})}{}_i \to R_k$，则存在影响者删失机制。2. *e*-共谋者：对于给定的 $R_k \in \mathbf{R}^{\text{aff}(i)}$ 和 $R_j \notin \mathbf{R}^{\text{aff}(k)}$，若 $Z^{(1,\mathbf{r}_{\text{aff}(i)}=\mathbf{1})}{}_i \to R_j \leftarrow R_k$，则在 R_j 处存在一个（纠缠的或）e-共谋者。3. *e*-共谋路径：对于给定的 $R_k \in \mathbf{R}^{\text{aff}(i)}$，若 $Z^{(1,\mathbf{r}_{\text{aff}(i)}=\mathbf{1})}{}_i$ 且 R_k 之间通过至少一条 [8] （不经过观测代理变量的）碰撞路径相连，则定义这对 $(Z^{(1,\mathbf{r}_{\text{aff}(i)}=\mathbf{1})}{}_i, R_k)$ 之间存在一条（纠缠的或）e-共谋路径。

定理 4。在具有缺失干扰的缺失数据ADMG $\mathcal{G}$中，在假设 1下，若且唯若该祖先有向混合图中对任意 $R \in \mathbf{R}$ 都不存在 e-共谋路径，则完全可观测分布 $P(\tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})}, \mathbf{R})$ 可识别。此外，若 $\mathcal{G}$ 是缺失数据有向无环图，则若且唯若不存在 e-共谋者且不存在 e-自我审查，完全可观测分布才可识别。识别泛函为

$$p(\tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})}, \mathbf{R}) = p(\tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})}) \times \underbrace{p(\mathbf{R} \mid \tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})})}_{g(p(\mathbf{R},\mathbf{Z}))} = \frac{p(\tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})}, \mathbf{R}=1)}{\underbrace{p(\mathbf{R}=1 \mid \tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})})}_{g(p(\mathbf{R},\mathbf{Z}))|_{\mathbf{R}=1}}} \times \underbrace{p(\mathbf{R} \mid \tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})})}_{g(p(\mathbf{R},\mathbf{Z}))}. \tag{9}$$

并且使用下文给出的 OR 参数化来识别缺失机制 $p(\mathbf{R} \mid \tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})})$：

$$p(\mathbf{R}|\tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})}) = \frac{1}{\sigma} \times \prod_{k=1}^{K} p(R_k|\mathbf{R}_{-k}=1, \tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})}) \times \prod_{k=2}^{K} \text{OR}(R_k, R_{\prec k}|R_{\succ k}=1, \tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})})$$

[7]这些结构中的“e”代表“纠缠的”。[8]若路径上的每个顶点 V_k 都是碰撞点（即具有以下形式之一：$\to V_k \leftarrow$、$\leftrightarrow V_k \leftrightarrow$、$\to V_k \leftrightarrow$或$\leftrightarrow V_k \leftarrow$），则顶点 V_i 与 V_j 之间的路径称为碰撞路径。

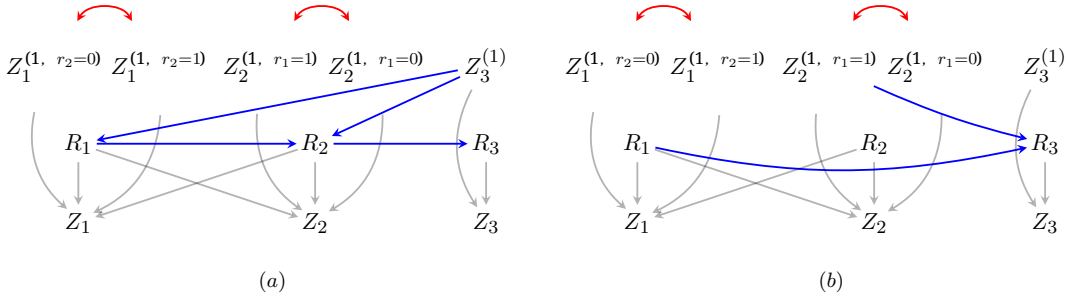


图 7：用于说明定理 4 的两个示例。(a) 中的无环有向混合图没有 e-共谋路径。(b) 中的无环有向混合图包含 e-共谋路径 $Z^{(1,r_1=1}{}_{)\,2} \rightarrow R_3 \leftarrow R_1$。

其中，符号表示法和优势比与第4.1节一致。

证明： 该证明扩展了在不存在共谋路径的情况下，对无环有向混合图中缺失数据的完整分布进行可靠且完备识别的证明，见 [12]。证明见附录。□

我们通过两个不同的示例说明定理 4：一个示例中完整可观测性分布可被识别，另一个示例中则违反了图形判据，因此该分布无法被识别。

在图 7(a) 所示的祖先有向混合图中，不存在 e-共谋路径。因此，根据定理 4，完整可观测性分布 $P(Z^{(1,r_2=2}{}_{)\,1}, Z^{(1,r_1=1}{}_{)\,2}, Z_3^{(1)}, R_1, R_2, R_3)$ 是可识别的。我们将在下一段中说明具体方法。

R_3 的倾向评分可以直接从观测数据中识别：$p(R_3|\text{pa}(R_3)) = p(R_3|R_2)$。但对于 R_2、R_1 的倾向评分，情况并非如此。对于 R_1，倾向评分为 $p(R_1|Z_3^{(1)})$。我们看到 $R_1 \perp\!\!\!\perp R_3|Z_3^{(1)}$、$R_2$，但如果没有 R_2，独立性就不成立。也就是说，我们可以识别 $p(R_1|Z_3^{(1)}, R_2, R_3 = 1)$。对于 R_2，倾向评分为 $p(R_2|Z_3^{(1)}, R_1)$。这里，我们无法在条件化符号后加入 R_3，因为条件独立性不成立。不过，我们可以固定 $R_3 = 1$；这一操作会消除 R_3 对 R_2 的依赖，并得到一个满足 $R_2 \perp\!\!\!\perp R_3|Z_3^{(1)}$、$R_1$ 的世界（更多详情见附录）。但我们可以使用优势比参数化来识别联合评分 $p(R_1, R_2|Z_3^{(1)})$。

$$p(R_1, R_2|Z_3^{(1)}) = \frac{1}{Z} \times p(R_2|R_1 = 1, Z_3^{(1)}) p(R_1|R_2 = 1, Z_3^{(1)}) \times \text{OR}(R_2, R_1|Z_3^{(1)})$$

在此，$p(R_2|R_1 = 1,\ Z_3^{(1)})$ 和 $p(R_1|R_2 = 1,\ Z_3^{(1)})$ 可以按照上一段所述进行识别。此外，我们可以将优势比展开如下：

$$\begin{aligned}\text{OR}(R_2, R_1|Z_3^{(1)}) &= \frac{p(R_1 = r_1|R_2 = r_2, Z_3^{(1)})}{p(R_1 = 1|R_2 = r_2, Z_3^{(1)})} \times \frac{p(R_1 = 1|R_2 = 1, Z_3^{(1)})}{p(R_1 = r_1|R_2 = 1, Z_3^{(1)})} \\ &= \frac{p(R_1 = r_1|R_2 = r_2, Z_3^{(1)}, R_3 = 1)}{p(R_1 = 1|R_2 = r_2, Z_3^{(1)}, R_3 = 1)} \times \frac{p(R_1 = 1|R_2 = 1, Z_3^{(1)}, R_3 = 1)}{p(R_1 = r_1|R_2 = 1, Z_3^{(1)}, R_3 = 1)}\end{aligned}$$

它也得到了识别。归一化项 Z 也得到了识别，因为条件部分和优势比部分都分别得到了识别。这意味着 $p(R_1,\ R_2|Z_3^{(1)})$ 得到了识别。

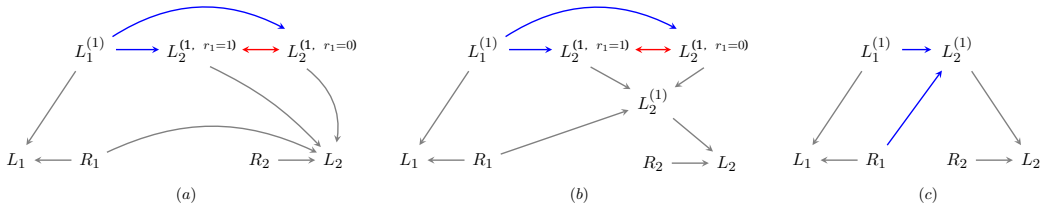


图 8：用于说明独立同分布设置下缺失干扰作用的图。

根据公式 9，完整可观测性分布得到了识别，因为所有部分都分别是观测数据的函数。

然而，由于共谋路径 $Z^{(1,r_1=1)}{}_2 \quad \rightarrow R_3 \leftarrow R_1$，图 7(b) 中的完整可观测性分布并未得到识别。直观地说，对于 R 的所有取值，都应当识别缺失机制 $p(\mathrm{R}\ |\mathrm{pa(R)})$。将缺失机制写为 $p(R_1) \times p(R_2) \times p(R_3\ |R_1, Z^{(1,r_1=1)}\ 2\ )$，我们会发现，除非能够将$R_1$ 和 R_2 都设为值 1，并且根据一致性，将反事实 $Z^{(1,r_1=1)}\ 2$ 替换为观测代理变量 Z_2，否则无法识别最后一项。在联合分布中这样做意味着，对于将 R_1 或 R_2 设为 0 的 R 取值，我们无法识别该量。关于不可识别性的严格论证，需要提供两个缺失数据无环有向混合图：它们的完整可观测性分布不一致，但观测数据分布一致。例如，可以在二元模型中，对刻画完整可观测性分布所需的参数与刻画观测分布所需的参数进行计数，并证明观测分布的参数更少，因此不可能将其唯一映射回完整可观测性分布。关于含有 e-共谋路径的示例中的参数计数，详细说明已推迟至附录；按照后续定理 4 证明中的完备性部分所采用的参数计数过程，需要掌握二元变量无环有向混合图的嵌套马尔可夫模型的莫比乌斯参数化，这一点我们也将在附录中讨论。

5. 缺失干扰模型对独立同分布设置的启示。在目前所述内容中，纠缠缺失的图模型被视为表示一组存在依赖或干扰的单位。可是，同样的图结构也可以被视为表示独立同分布设置下单个单位的多变量观测，例如，单个单位具有多变量处理和／或结局测量值的设置。

如我们所示，目标分布依赖和缺失过程依赖可以通过现有的独立同分布缺失数据方法加以处理。然而，据我们所知，既有文献从未讨论过独立同分布设置下缺失数据中与缺失干扰纠缠相对应的情形。事实上，经典缺失数据图方法中普遍采用的边限制通常要求不存在缺失指示变量作为任何反事实变量或完全观测变量的原因／父节点（第2.3节中的限制（ii））。放宽这一假设意味着，缺失指示变量的变化可能导致底层完整数据分布发生变化。我们提出的缺失干扰纠缠框架可以很容易地加以调整，从而放宽独立同分布设置中施加的边限制。

例如，假设 L_1，L_2 分别是在第一次和第二次随访就诊时获得的测量值（i 中的 L_i 是单个患者的索引变量）。在某些情况下，患者是否参加首次随访就诊可能会影响其未来健康状况；或许是通过

心理上的自我赋能或自我鼓励的话语来影响治疗依从性。这可以通过以下形式的反事实来表示 $L^{(r_2=1,r_1)}{}_2$，或者简写为 $L^{(1,r_1)}{}_2$，其含义是：如果患者参加了首次就诊（$r_1=1$），那么在第二次就诊时获得的测量值会是多少；或者如果患者没有参加首次就诊（($r_1=0$）。这个例子类似于假设 L_1 是某位患者的测量值，而 L_2 是另一位患者首次就诊时的测量值（i 在 L_i 中，索引个体对应同一个变量）。假设两位患者之间存在非对称关系，即只有第二位患者的测量值取决于第一位患者是否参加就诊，反之则不然，我们可以用图8(a)所示的祖先有向混合图表示这种关系。该祖先有向混合图对应的完整数据分布中，有一个反事实版本（由索引），而有两个反事实版本（分别由，$R_1=0$和 $R_2=1$，$R_1=1$索引）。

与非独立同分布情形中的缺失干扰类似，代理变量 L_2 可以定义为 $R_1, R_2, L^{(1,r_1=0)}{}_2$ 和 $L^{(1,r_1=1)}{}_2$ 的确定性函数，如下所示：

$$L_2 \leftarrow \begin{cases} (1-r_1)\, L_2^{(1,r_1=0)} + r_1\, L_2^{(1,r_1=1)} & \text{if } r_2=1, \\ ? & \text{if } r_2=0. \end{cases}$$

我们还可以先将 L_2 的单索引反事实定义为 $L^{(r_2=1)}{}_2$ 或简称 $L_2^{(1)}$，并将其定义为 $R_1, L^{(1,r_1=1)}{}_2, L^{(1,r_1=0)}{}_2$ 的确定性函数；然后再将代理变量 L_2 仅定义为 $R_2, L_2^{(1)}$ 的确定性函数：

$$L_2^{(1)} = (1-R_1) \times L_2^{(1,r_1=0)} + R_1 \times L_2^{(1,r_1=1)},$$

$$L_2 \leftarrow \begin{cases} L_2^{(1)} & \text{if } r_2=1, \\ ? & \text{if } r_2=0. \end{cases}$$

图 8(b) 中的祖先有向混合图对应于包含L_2的三个反事实版本的完整数据分布。灰色边展示了上述一致性关系。

图 8(a, b) 与经典独立同分布缺失数据有向无环图的区别在于存在多重索引反事实，这表明，例如，在给定变量时间顺序的情况下，顺序靠后的变量可能取决于较早的变量是否被测量，而无论样本实现之间是否相互依赖。将图 8(b) 中的多重索引反事实投影掉后，得到图 8(c) 中的有向无环图，其中 R_1 指向 $L_2^{(1)}$，这违反了经典缺失数据图中的边限制。

在上述独立同分布示例中，由于对 R_1 的赋值不一致，目标分布 $p(L^{(r_1=1)}{}_1, L^{(r_2=1,r_1=0)}{}_2)$ 无法作为观测数据分布的函数加以识别，除非施加严格的跨世界假设 [9]。例如，我们可以假设，$L^{(r_1=1)}{}_1$ 对 $L^{(r_2=1)}{}_2$ 的效应完全由一个不存在缺失性的变量所中介，该变量记为 X。那么，$p(L^{(r_1=1)}{}_1, L^{(r_1=0,r_2=1)}{}_2)$ 可以如下识别：

$$\begin{aligned} p(L_1^{(r_1=1)}, L_2^{(r_2=1,r_1=0)}) &= \sum_x p(L_1^{(r_1=1)} \mid x) \times p(L_2^{(r_2=1,r_1=0)} \mid x, L_1^{(r_1=1)} = l_1^{(r_1=1)}) \times p(x) \\ &= \sum_x p(L_1 \mid x, R_1=1) \times p(L_2 \mid x, R_1=0, R_2=1) \times p(x). \end{aligned}$$

为避免此类复杂情况，既有关于缺失数据图模型的文献排除了从$\mathbf{R}$ 指向反事实的边。在此边限制下，我们假设 $L_j^{(r_k,r_j=1} = L_j^{(r_j=1)})$，$\forall R_j$，$R_k \in R$. 如上所示，利用我们所建立的缺失干扰纠缠框架，可以放宽这一假设。

[9] 关于跨世界假设的讨论可见于中介分析文献。例如，参见 [85]

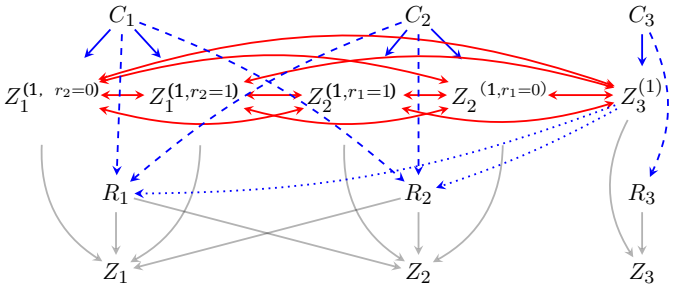


图 9：我们实验所用合成数据的生成模型。

6. 模拟。我们生成合成数据，并尝试在目标可识别的纠缠缺失性情形下恢复真实值。具体而言，我们关注存在缺失干扰（$\mathcal{E}_3$）的情形；关于不存在缺失干扰的情形，既有文献已进行了讨论 [16]。

数据根据图9进行了模拟。网络中有三个个体，个体 1 和 2 是邻居。当图仅包含实线边时，我们处于完全随机缺失（MCAR）情形。加入蓝色虚线边（$C_1 \to R_1$, $C_1 \to R_2$, $C_2 \to R_1$,$C_2 \to R_2$, $C_3 \to R_3$）后，得到随机缺失（MAR）情形。最后，加入所有边（再添加 $Z_3^{(1)} \to R_2, Z_3^{(1)} \to R_1$）后，得到非随机缺失（MNAR）情形。

与生成模型相关的参数如下： $p(C_1)$ $p(C_2)$ $p(C_3)$，均设为单变量正态分布；反事实变量 $Z^{(1,r_2=0)}{}_1$, $Z^{(1,r_2=1)}{}_1$ 、 $Z^{(1,r_1=0)}{}_3$ 、 $Z^{(1,r_1=1)}{}_3$ 和 $Z_3^{(1)}$ 服从多变量正态分布。多变量正态分布的均值是图中相应父节点 C 的线性函数，且协方差矩阵中没有零元素；这确保了所有反事实变量彼此相关。缺失指示变量 R_i 根据相应情景生成，采用二项分布，其中二项概率完全随机选择（完全随机缺失，MCAR），或设为 R_i 的父节点的线性函数经逻辑 sigmoid 函数变换后的值（随机缺失，MAR；非随机缺失，MNAR）。观测代理变量 Z_i 是反事实变量和缺失指示变量的确定性函数，具体由一致性假设给出。

我们关注识别以下泛函： $\mathbb{E}[Z^{(1,r_2=0)}{}_1]$, $\mathbb{E}[Z^{(1,r_2=1)}{}_1]$, $\mathbb{E}[Z^{(1,r_1=0)}{}_2]$, $\mathbb{E}[Z^{(1,r_1=1)}{}_2]$,在完全随机缺失、随机缺失和非随机缺失条件下。在给定的图中，由于不存在 e-共谋者，即使在非随机缺失情况下，我们也应该能够识别完整可观测性分布（或其边缘分布）。我们使用以下形式的估计方程： $\mathbb{E}[\mathbb{I}(\mathbf{R}=\mathbf{1})\pi(\mathbf{C},\tilde{\mathbf{Z}})h(\tilde{\mathbf{Z}})]$

其中， $\pi(\mathbf{C},\mathbf{Z})$ 是由倾向评分构成的逆权重函数， $h(.)$ 是我们关注的泛函。该估计量使用回归模型估计权重，并通过经验数据计算期望。

我们希望将得到的调整后估计值，与一位不采用纠缠缺失思路的研究者所得到的估计值进行比较，以充分说明这种框架为何不可或缺。首先，当这样一位不了解单位间各种相互作用的研究者在数据集中发现缺失数据时，很可能会采用朴素的缺失数据方法来估计反事实，并可能假设数据遵循独立同分布的随机缺失机制。这意味着，他们只能估计 $Z_1^{(1)}$ 和 $Z_2^{(1)}$；由于单位1和单位2各有两个 反事实，这些估计值无法可靠地反映真实情况。为复现此类分析，我们在独立同分布的随机缺失假设下估计 $Z_1^{(1)}$ 和 $Z_2^{(1)}$。许多估计量

表 1基于50次自助法试验得到的反事实估计值及其分位数 $q_{0.05}$ 和 $q_{0.95}$。不了解纠缠关系的研究者只能假设缺失数据情景为独立同分布，并很可能在随机缺失假设下估计Z（1）1 和 Z（1）2，如最后两行所示；此处使用增广逆概率加权估计量完成估计。这些值无法以任何可靠的方式对应于前四行所记录的四个潜在反事实；这些反事实是在了解纠缠关系及相应所需调整的基础上估计得到的。

估计值	Mean	$q_{0.05}$	$q_{0.95}$
$Z_1^{(1,\ r_2=1)}$	-2.942	-3.009	-2.889
$Z_1^{(1,\ r_2=0)}$	3.138	3.108	3.164
$Z_2^{(1,\ r_1=1)}$	10.816	10.773	10.856
$Z_2^{(1,\ r_1=0)}$	-4.333	-4.351	-4.313
$Z_1^{(1)}$	1.497	1.456	1.536
$Z_2^{(1)}$	2.252	2.138	2.373

存在适用于此类情形的方法，包括 g-公式和基于逆概率权重（逆概率加权）的估计量 [86]。在本实验中，我们使用由 $\frac{1}{N}\sum_{n=1}^{N}\left[\frac{R_i}{R_i^{`}}{}_{n}Z_{i,}\ p\ {}_{n}|C_{i\ n}^{n}\ (\ .\ .\ )\ \_\ \{R_{i,n}-p\ (\ R_{i\ n}|C_{i\ n}\ \}\ .\ .\ )\ (\ .\ .\ )\ \mathbb{E}[Z_i|C_{i,n}]\right]$ 给出的双重稳健增强型逆概率加权估计量，其中 i 表示个体，n是样本。有关这些估计量的更多信息，请参阅 [86] 。

共从网络中生成了 $50k$ 个样本，并通过 50 次自助法试验进行估计。所得估计值见表 1。前四行展示了对反事实进行适当调整后的估计值，后两行则展示了完全忽略个体间依赖、但采用随机缺失假设处理缺失性的估计值。

为展示调整后的估计值的表现，我们在两种不同情景下比较估计值与真实值（根据模型参数获得）之间的偏差：一种是适当针对网络结构和纠缠进行调整，另一种是不进行调整，分别考察随机缺失（MAR）和非随机缺失（MNAR）情形。[10] 后者是指直接对不同缺失模式采用条件均值。未经调整、基于完整行得到的估计值会产生偏差，如图 10 所示。调整后得到的估计值在横轴上以星号 $(*)$ 标示。

正如我们在此指出的，如果研究者不谨慎地同时考虑依赖和缺失性以及二者相互作用的确切方式，其研究方法可能会在诸多方面产生偏差；我们这里只是说明了其中一种可能的情况。

[10]在完全随机缺失（MCAR）情形下，两种不同方法产生的偏差应当相同，因为我们可以将数据集中的完整行视为一个无偏的代表性数据集。

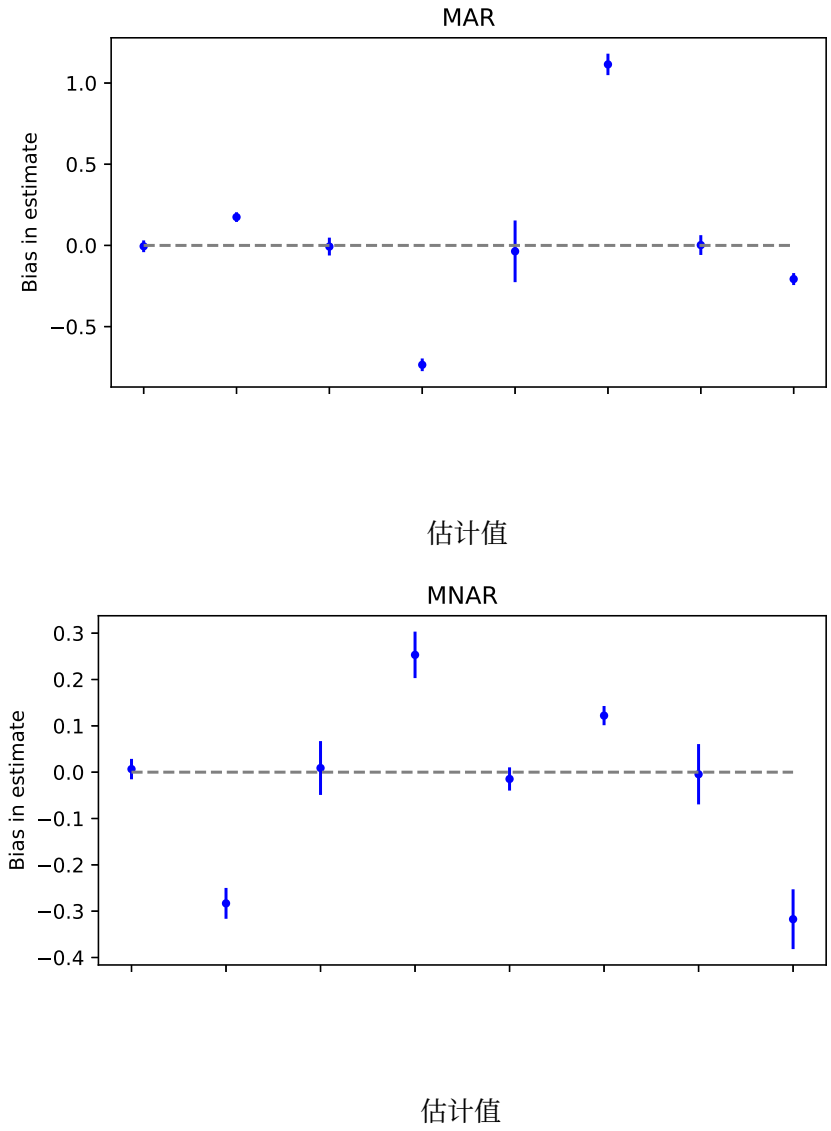


图 10：目标（横轴所示）的自助抽样估计值中记录的偏差：随机缺失情形（左）和非随机缺失情形（右）。我们将调整后的逆概率加权估计值（以星号（$*$） 表示，位于横轴上）与未调整的估计值进行比较；后者通过忽略数据所依赖的网络缺失结构得到。误差条表示 50 个自助抽样样本中的分位数 $q_{0.05}$ 和 $q_{0.95}$。

7. 讨论。本文提出了一个图形框架，用于建模数据依赖与缺失数据同时存在所导致的纠缠。在存在部分干扰的非独立同分布设置中，纠缠可能源于不同个体的目标分布变量之间的依赖（目标分布依赖）、一个个体的缺失指示变量对其他个体的目标分布变量的依赖（缺失指示变量依赖），或一个个体的观测事实变量对其他个体缺失指示变量的依赖（缺失干扰）。我们以数学方式刻画了这三种纠缠情形，并针对每种情形给出了完整分布识别理论的可靠性和完备性。我们注意到，纠缠缺失模型的识别中仍有若干尚未解决的问题，这些问题在本文中没有涉及。例如，虽然我们关注的是完整可观测性分布，但还有许多模型即使在完整（可观测性）分布无法识别时，特定目标或参数（如因果效应）仍可能可以识别。即使在独立同分布的缺失数据模型中，也存在可识别的目标分布，而相应的完整分布却无法识别；事实上，目标分布识别的完整刻画仍是一个开放问题。最后，我们证明，我们的缺失干扰模型扩展了独立同分布图形缺失数据方法，使其能够处理缺失指示变量可能成为反事实变量或完全观测随机变量原因的情形。这种现象在现实场景中（尤其是纵向数据中）十分常见，但此前缺失数据文献尚未对此进行研究。

附录 A：嵌套马尔可夫因子分解

相对于祖先有向混合图 $\mathcal{G}$(V)， p(V) 的嵌套马尔可夫分解通过条件分布来定义，这些条件分布称为核；此外还要用到与之关联的条件祖先有向混合图（称为条件化无向无环混合图，CADMG）。这些图分别由 p(V) 和 G(V) 经由反复应用固定算子 [62] 推导而来。条件化无向无环混合图 $\mathcal{G}$(V,W) 是一种祖先有向混合图，其节点可划分为随机变量 V 和固定变量 W；并且规定只有出边可以与 W 中的变量相邻。更一般地，也可以针对分布 p(V|W) 和条件化无向无环混合图 $\mathcal{G}$(V,W) 定义嵌套马尔可夫分解，并将核与映射联系起来；这些映射由 p(V|W) 推导而来，而相应的条件化无向无环混合图则由 $\mathcal{G}$(V,W) 经由固定操作推导而来。

核：核 $q_{\mathbf{V}}$(V|W) 是一种映射，它将 W 中的取值映射为关于 V[87] 的归一化密度。条件分布是核的一个常见例子，因为 $\sum_{\mathbf{v}\in\mathbf{V}} q_{\mathbf{V}}(\mathbf{v}|\mathbf{w}) = 1$。核中的条件化和边缘化按通常方式定义：对于 $\mathbf{A}\subseteq\mathbf{V}$，$q_{\mathbf{V}}(\mathbf{A}|\mathbf{W}) \equiv \sum_{\mathbf{V}\setminus\mathbf{A}} q_{\mathbf{V}}(\mathbf{V}|\mathbf{W})$，且 $q_{\mathbf{V}}(\mathbf{V}\setminus\mathbf{A}|\mathbf{A}\cup\mathbf{W}) \equiv \frac{q_{\mathbf{V}}(\mathbf{V}|\mathbf{W})}{q_{\mathbf{V}}(\mathbf{A}|\mathbf{W})}$。

可固定性与固定算子： 在条件化无向无环混合图 $\mathcal{G}$ 中，变量 $V\in\mathbf{V}$ 可固定，当 $\mathrm{de}_{\mathcal{G}}(V)\cap \mathrm{des}_{\mathcal{G}}(V) = \emptyset$。换言之，对于任意 $B\in\mathbf{V}\setminus\{V\}$，若路径 $V\leftrightarrow\ldots\leftrightarrow B$ 和 $V\rightarrow\ldots\rightarrow B$ 在 $\mathcal{G}$ 中并非同时存在，则 V 可固定。
我们为图定义固定算子 $\phi_V(\mathcal{G})$，为核定义固定算子 $\phi_V(q;\mathcal{G})$。给定一个 CADMG $\mathcal{G}$(V, W)，其中有一个可固定的顶点 $V\in$V，$\phi_V(\mathcal{G}$(V, W)) 会生成一个新的 CADMG$\mathcal{G}$(V $\setminus\{V\}$, W $\cup\{V\}$)：它由 $\mathcal{G}$(V, W) 将 V从 V 移至 W，并移除所有箭头指向 V的边而得到。给定一个核 $q_{\mathbf{V}}$(V|W) 和一个 CADMG $\mathcal{G}$(V, W)，算子 $\phi_V(q_{\mathbf{V}}$(V|W) $\mathcal{G}$(V, W)) 会生成一个新核：

$$q_{\mathbf{V}\setminus\{V\}}(\mathbf{V}\setminus\{V\}|\mathbf{W}\cup\{V\}) \equiv \frac{q_{\mathbf{V}}(\mathbf{V}|\mathbf{W})}{q_{\mathbf{V}}(V|\,\mathrm{mb}_{\mathcal{G}}(V))}$$

固定序列： 如果序列 $\langle V_1,\ldots,V_k\rangle$ 中的每个顶点都能在 $\mathcal{G}(\mathbf{V},\ \mathbf{W})$ 中固定，那么该序列称为在其中有效；V_1 在$\mathcal{G}(\mathbf{V},\ \mathbf{W})$ 中可固定，V_2 在 $\phi_{V_1}(\mathcal{G}(\mathbf{V},\ \mathbf{W}))$ 中可固定，依此类推。对于同一集合 $\mathbf{S}\subseteq\mathbf{V}$，若任意两个序列 σ_1，σ_2均可在 $\mathcal{G}$ 中固定，则它们会得到相同的 CADMG。图固定算子可扩展到集合 $\mathbf{S}$：$\phi_{\mathbf{S}}(\mathcal{G})$。该算子定义为对集合 $\mathbf{S}$ 的任意有效序列 σ 依次应用顶点固定操作。

给定一个序列 $\sigma_{\mathbf{S}}$，定义 $\eta(\sigma_{\mathbf{S}})$ 为 $\sigma_{\mathbf{S}}$ 中的第一个元素，定义 $\tau(\sigma_{\mathbf{S}})$ 为 $\sigma_{\mathbf{S}}$ 中除第一个元素外包含所有元素的子序列。给定一个在 S 上、在 $\mathcal{G}$(V, W) 中有效的元素序列 $\sigma_{\mathbf{S}}$，核固定算子 $\phi_{\sigma_{\mathbf{S}}}(q_{\mathbf{V}}$(V|W) $\mathcal{G}$(V, W)) 定义为：若 $\sigma_{\mathbf{S}}$ 为空序列，则等于 $q_{\mathbf{V}}$(V|W)；否则等于 $\phi_{\tau(\sigma_{\mathbf{S}})}(\phi_{\eta(\sigma_{\mathbf{S}})}(q_{\mathbf{V}}$(V|W); $\mathcal{G}$(V, W)),$\phi_{\eta(\sigma_{\mathbf{S}})}(\mathcal{G}$(V, W)))。

可达性： 给定一个条件化无向无环混合图 $\mathcal{G}(\mathbf{V},\ \mathbf{W})$，若集合 $\mathbf{R}\subseteq\mathbf{V}$ 存在一个在 $\mathcal{G}(\mathbf{V},\ \mathbf{W})$ 中有效的、针对 $\mathbf{V}\setminus\mathbf{R}$ 的序列，则称其为可达的 。换言之，若 $\mathbf{S}$ 在 $\mathcal{G}$ 中可固定，则 $\mathbf{V}\setminus\mathbf{S}$ 可达。

内在集： 若集合 $\mathbf{R}$ 在 $\mathcal{G}(\mathbf{V},\ \mathbf{W})$ 中可达，则当 $\phi_{\mathbf{V}\setminus\mathbf{R}}(\mathcal{G})$ 仅包含一个区（即集合 $\mathbf{R}$ 本身）时，称 R 为 $\mathcal{G}(\mathbf{V},\ \mathbf{W})$ 中的内在集 。CADMG $\mathcal{G}$ 中内在集的集合记作 $\mathcal{I}(\mathcal{G})$。

嵌套马尔可夫分解： 若存在一组如下形式的核，则称分布 $p(\mathbf{V}|\mathbf{W})$ 满足关于条件化无向无环混合图 $\mathcal{G}(\mathbf{V},\mathbf{W})$ 的嵌套马尔可夫分解

$\{q_{\mathbf{S}}(\mathbf{S}|\mathrm{pa}_{\mathcal{G}}(\mathbf{S})):\ \mathbf{S}\in\mathcal{I}(\mathcal{G})\}$ 使得对于每个可达集$\mathbf{R}$ 的有效序列 $\sigma_{\mathbf{R}}$，在$\mathcal{G}$中都有：

$$\phi_{\sigma_{\mathbf{R}}}(p(\mathbf{V}|\mathbf{W});\mathcal{G}(\mathbf{V},\mathbf{W}))=\prod_{\mathbf{D}\in\mathcal{D}(\phi_{\mathbf{R}}(\mathcal{G}(\mathbf{V},\mathbf{W})))}q_{\mathbf{D}}(\mathbf{D}|\,\mathrm{pa}^{\mathrm{s}}{}_{\mathcal{G}}(\mathbf{D}))$$

如果一个分布遵循该因子分解，那么对于任意可达的 R，在 p(V|W) 上应用于 R 的任意两个有效序列都会得到相同的核 $q_{\mathbf{R}}$(R|V \R)。因此，核固定可以像图固定一样定义在集合上。在这种情况下，对于每个 D $\in\mathcal{I}(\mathcal{G})$， $q_{\mathbf{D}}$(D|pa$^{\mathrm{s}}{}_{\mathcal{G}}$(D)) $\equiv\phi_{\mathbf{V}\setminus\mathbf{D}}$($p$(V|W); $\mathcal{G}$(V,W))。

p区划分解或田分解$(\mathbf{V}|\mathbf{W})$ 由嵌套分解得到：

$$\begin{aligned}p(\mathbf{V}|\mathbf{W})&=\prod_{\mathbf{D}\in\mathcal{D}(\mathcal{G}(\mathbf{V},\mathbf{W}))}q_{\mathbf{D}}(\mathbf{D}|\,\mathrm{pa}^{\mathrm{s}}{}_{\mathcal{G}}(\mathbf{D}))\\&=\prod_{\mathbf{D}\in\mathcal{D}(\mathcal{G}(\mathbf{V},\mathbf{W}))}\left(\prod_{D\in\mathbf{D}}p(D\mid\mathrm{pre}_{\prec}(D))\right),\end{aligned}$$

其中，前驱$\prec(D)$ 是根据拓扑全序$\prec$确定的 D 的前驱集合。每个因子 $\prod_{D\in\mathbf{D}}p(D\mid$前驱$\prec(D))$ 在嵌套分解下仅是$\mathbf{D}\cup\mathrm{pa}_{\mathcal{G}}(\mathbf{D})$ 的函数。

中的一个重要结果 [62] 指出，若 $p(\mathbf{V}\cup\mathbf{H}|\mathbf{W})$ 遵循某个 CDAG$\mathcal{G}(\mathbf{V}\cup\mathbf{H},\mathbf{W})$的因子分解，那么 $p(\mathbf{V}|\mathbf{W})$ 遵循潜在投影 CADMG$\mathcal{G}(\mathbf{V},\mathbf{W})$的嵌套分解。

识别：并非每个干预分布 $p(\mathbf{Y}(\mathbf{a}))$ 都能在隐变量因果模型中识别。然而，从 $p(\mathbf{V}|\mathbf{W})$ 识别出的每个 $p(\mathbf{Y}(\mathbf{a})|\mathbf{W})$ 都可以表示为如下的修正嵌套分解：

$$\begin{aligned}&p(\mathbf{Y}(\mathbf{a})|\mathbf{W})\\&\quad=\sum_{\mathbf{Y}^*\setminus\mathbf{Y}}\prod_{\mathbf{D}\in\mathcal{D}(\mathcal{G}_{\mathbf{Y}^*})}p(\mathbf{D}|\,\mathrm{do}(\mathrm{pa}^{\mathrm{s}}{}_{\mathcal{G}}(\mathbf{D})))|_{\mathbf{A}=\mathbf{a}}\\&\quad=\sum_{\mathbf{Y}^*\setminus\mathbf{Y}}\prod_{\mathbf{D}\in\mathcal{D}(\mathcal{G}_{\mathbf{Y}^*})}\phi_{\mathbf{V}\setminus\mathbf{D}}(p(\mathbf{V}|\mathbf{W});\mathcal{G}(\mathbf{V},\mathbf{W}))|_{\mathbf{A}=\mathbf{a}}\end{aligned}$$

其中$\mathbf{Y}^*\equiv$是$\mathcal{G}(\mathbf{V}(\mathbf{a}),\mathbf{W})(\mathbf{Y})\setminus\mathbf{a}$。也就是说，只有当 $p(\mathbf{Y}(\mathbf{a})|\mathbf{W})$ 可以表示为一个因子分解时，它才能被识别；其中每一部分都对应于 $\mathcal{G}(\mathbf{V},\mathbf{W})$ 中某个内在集合所关联的核。此外，该因子分解中的任何部分都不包含作为随机变量的$\mathbf{A}$ 元素。

嵌套马尔可夫模型的二元参数化：根据嵌套分解，给定父节点的内在集合构成嵌套马尔可夫模型的原子单位。利用这一观察，[88] 给出了离散嵌套马尔可夫模型的一种光滑参数化。我们简要说明如何推导二元嵌套马尔可夫模型的莫比乌斯参数。

对于每个区 D $\in\mathcal{D}(\mathcal{G})$，考虑 S $\subseteq$ S 的所有可能子集。若 S 在 $\phi_{\mathbf{V}\setminus\mathbf{S}}(\mathcal{G})$ 中是内在的，则将内在集的头部 H 定义为 S 中在 $\phi_{\mathbf{V}\setminus\mathbf{S}}(\mathcal{G})$ 中没有子节点的所有顶点，并将尾部 T 定义为 CADMG $\phi_{\mathbf{V}\setminus\mathbf{S}}(\mathcal{G})$ 中头部的所有父节点，但不包括头部本身。形式上，H $\equiv\{V\in\mathrm{S}|\mathrm{ch}_{\phi_{\mathbf{V}\setminus\mathbf{S}}(\mathcal{G})}(V)=\emptyset\}$ 且 $T\equiv\mathrm{pa}_{\phi_{\mathbf{V}\setminus\mathbf{S}}(\mathcal{G})}(\mathrm{H})\setminus\mathrm{H}$。相应的

这组内在头尾对的莫比乌斯参数可对核进行参数化 $q_{\mathbf{S}}(\mathbf{H}=0|\mathbf{T})$；也就是说，核中内在集 $\mathbf{S}$ 之外的所有变量都固定，并且在给定尾部的条件下，将头部的所有元素设为零。请注意，这些参数通常在变分上相互依赖（与普通有向无环图模型中变分独立的情况相反），因为这些参数集中的头部和尾部可能重叠。任何查询的联合密度$p(\mathbf{V}=\mathbf{v})$ 都可以通过莫比乌斯反演公式得到；详见 [87, 88]。我们将 $q_{\mathbf{S}}(\mathbf{H}=\mathbf{0}|\mathbf{T})$ 简记为 $q(\mathbf{H}=\mathbf{0}|\mathbf{T})$，因为通常可以清楚地判断，与给定内在集对应的核中哪些变量仍然是随机的。

缺失数据模型的二元参数化： 我们使用前文所述的参数化方法，计算对缺失数据ADMG的完整可观测性分布及其对应观测分布进行参数化所需的参数数量。随后据此推理：如果完整可观测性分布中的参数数量多于观测分布中的参数数量，就不可能建立从观测分布到完整分布的映射。这进而意味着该完整可观测性分布无法识别。在完整可观测性分布中，可以忽略确定性因子，即给定父节点的代理变量，因为这些事件的概率始终为 1。然而，在计算观测分布的参数数量时，我们谨慎地将反事实视为未观测，并据此得到相应的 ADMG。随后以类似前文的方式推导莫比乌斯参数，但还需满足以下约束：如果 Z_i 出现在某个参数的头部中，而缺失指示变量 R_i 或 $\mathbf{R}^{\text{aff}(i)}$ 出现在其尾部中，那么核必须限定在 $R_i=1$ 且 $\mathbf{R}_{\text{aff}(i)}=\mathbf{1}$的情形。这是因为：（1）当 $R_i=0$时，头部取除 $Z_i=?$ 之外任何值的概率都确定为 0；（2）$\mathbf{R}^{\text{aff}(i)}$ 取不同于 1 的值的情形，与识别完整可观测性分布无关。这里考虑的是 $R_i=1$ 始终成立，但只有在识别完整可观测性分布时才考虑 $\mathbf{R}_{\text{aff}(i)}=\mathbf{1}$。其他情形则需采用不同的考量。

附录 B：证明

我们在此重述定理并概述其证明。

定理 *3* 在一个具有缺失干扰的缺失数据*ADMG* $\mathcal{G}$ 中，当以下两个条件之一满足时，由一组反事实 $Z' \equiv \bigcup_i \{Z^{(\mathbf{1},\mathbf{r}_{aff(i)})}\ i\},\ i \in \{1,\cdots,n\}$ 构成的有效单世界对象 $h(\tilde{Z};r)$ 是可识别的：*(1)* $R' \perp\!\!\!\perp O, Z$（完全随机缺失），或 *(2)* $R' \perp\!\!\!\perp Z|O$（随机缺失），其中 R' 指代所有缺失指示变量的集合 R，这些变量在 $h(Z;r)$ 中为反事实建立索引。对象 $h(Z;r)$ 是 $p(Z', R, O)$ 的函数，其识别泛函为：

$$p(\mathbf{Z}',\mathbf{R},\mathbf{O}) = p(\mathbf{Z}',\mathbf{O}) \times p(\mathbf{R}|\mathbf{O},\mathbf{Z}') = \frac{p(\mathbf{Z}',\mathbf{R}=\mathbf{r},\mathbf{O})}{p(\mathbf{R}=\mathbf{r}\mid\mathbf{O})} \times p(\mathbf{R}\mid\mathbf{O}) \tag{10}$$

其中，倾向评分通过对祖先有向混合图（或有向无环图）因子分解应用简单的 m-分离（或 d-分离）规则获得。

证明：$\mathbf{Z}'$ 是 h（$\mathbf{Z}$；$\mathbf{r}$）中的反事实集合，使得它们在世界$\mathbf{R}=\mathbf{r}$中构成一个有效的单世界反事实。此外，设所有为$\mathbf{Z}'$ 编入索引的缺失指示变量集合为$\mathbf{R}'$，并且当$\mathbf{R}=\mathbf{r}$时，$\mathbf{R}'=\mathbf{r}'$。我们感兴趣的是识别一个分布 P（$\mathbf{Z}'$）或由其得到的函数 h（$\mathbf{Z}$；$\mathbf{r}$）。这里的关键因素是，缺失指示变量为 0 的变量不会出现在 h（。）中。

在情形（1）中，我们可以写成 $P(\mathbf{Z}') = P(\mathbf{Z}' \mid \mathbf{R}'=\mathbf{r}')$，因为 $\mathbf{R}' \perp\!\!\!\perp \mathbf{Z}$ 且 $\mathbf{Z}' \subset \mathbf{Z}$。根据一致性，我们可以将所有反事实替换为相应的代理变量，从而识别该对象。

在情形（2）中，我们可以写成 $P(\mathbf{Z}') = P(\mathbf{Z}' \mid \mathbf{O}) \times P(\mathbf{O}) = P(\mathbf{Z}' \mid \mathbf{O}, \mathbf{R}' = \mathbf{r}') \times P(\mathbf{O})$，因为 $\tilde{\mathbf{R}}' \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{O}$ 且 $\mathbf{Z}' \subset \mathbf{Z}$。根据一致性，我们可以通过将所有反事实替换为相应的代理变量来识别该分布。

□

定理4 对于具有缺失干扰的缺失数据*ADMG* $\mathcal{G}$，在假设*1*下，完全可观测分布 $P(\tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})}, \mathbf{R})$ 可识别，当且仅当该*ADMG*中对任意 $R \in \mathbf{R}$ 都不存在 *e*-共谋路径。此外，若 $\mathcal{G}$ 是缺失数据 *DAG*，则完全可观测分布可识别，当且仅当不存在 *e*-共谋者且不存在 *e*-自我审查。识别泛函为

$$p(\tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})}, \mathbf{R}) = p(\tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})}) \times \underbrace{p(\mathbf{R} \mid \tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})})}_{g(p(\mathbf{R},\mathbf{Z}))} = \frac{p(\tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})}, \mathbf{R}=1)}{\underbrace{p(\mathbf{R}=1 \mid \tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})})}_{g(p(\mathbf{R},\mathbf{Z}))|_{\mathbf{R}=1}}} \times \underbrace{p(\mathbf{R} \mid \tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})})}_{g(p(\mathbf{R},\mathbf{Z}))}.$$

并且，缺失机制 $p(\mathbf{R} \mid \tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})})$ 可使用下文给出的 *OR* 参数化进行识别：

$$p(\mathbf{R}|\tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})}) = \frac{1}{\sigma} \times \prod_{k=1}^{K} p(R_k|R_{-k}=1, \tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})}) \times \prod_{k=2}^{K} OR(R_k, R_{\prec k}|R_{\succ k}=1, \tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})})$$

证明：

可靠性： 不存在 e-共谋路径即可实现识别。

在 $Z^{(1,\mathbf{r}_{\text{aff}(i)})\,i}$ 与 $R_k \in \mathbf{R}^{\text{aff}(i)}$ 之间不存在 e-共谋路径，这意味着 $Z^{(1,\mathbf{r}_{\text{aff}(i)})\,i} \notin \text{mb}_{\mathcal{G}}(R_k)$，其中

$$\begin{aligned}\text{mb}_{\mathcal{G}}(V) = \{&\text{pa}_{\mathcal{G}}(V), \text{dis}_{\mathcal{G}}(V), \text{pa}_{\mathcal{G}}(\text{dis}_{\mathcal{G}}(V)), \text{ch}_{\mathcal{G}}(V), \text{pa}_{\mathcal{G}}(\text{ch}_{\mathcal{G}}(V)),\\ &\text{dis}_{\mathcal{G}}(\text{ch}_{\mathcal{G}}(V)), \text{pa}_{\mathcal{G}}(\text{dis}_{\mathcal{G}}(\text{ch}_{\mathcal{G}}(V)))\}\end{aligned}$$

令 $\tilde{\mathbf{Z}}^{(\mathbf{r}_k=\mathbf{1})} = \{Z_i^{(1,\mathbf{r}_{\text{aff}(i)}=\mathbf{1})} : R_k \in \mathbf{R}_{\text{aff}(i)}\} \cup Z_k^{(1,\mathbf{r}_{\text{aff}(k)})}$。换言之，$\tilde{\mathbf{Z}}^{(\mathbf{r}_k=\mathbf{1})}$ 是对应于完全可观测性的所有反事实的集合，并由 R_k 索引（且受其影响）。

根据马尔可夫性质，我们有 $V \perp\!\!\!\perp \mathbf{V} \setminus \text{mb}_{\mathcal{G}}(V) | \text{mb}_{\mathcal{G}}(V)$。因此，不存在 e-共谋路径意味着以下假设：

$$R_k \perp\!\!\!\perp \{Z_i^{(1,\mathbf{r}_{\text{aff}(i)})} \in \tilde{\mathbf{Z}}^{(\mathbf{r}_k=\mathbf{1})}\} \mid \{R \setminus R_k, \tilde{\mathbf{Z}} \setminus Z_i^{(1,\mathbf{r}_{\text{aff}(i)}=\mathbf{1})}\} \quad \text{for } R_k \in \mathbf{R}$$

基于这些假设，我们可以使用 OR 参数化来识别 $p(\mathbf{R}|\tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})})$。证明与文献 [89] 中给出的无自我删失模型的识别证明以及文献 [12] 中对其的表示类似。由于我们假设 $\text{ch}_{\mathcal{G}}(\tilde{\mathbf{Z}}^{(\mathbf{r}\neq\mathbf{1})}) \cap \mathbf{R} = \emptyset$，因此只需识别 $p(\mathbf{R}|\tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})})$。

$$p(\mathbf{R}|\tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})}) = \frac{1}{\sigma} \times \prod_{k=1}^{K} p(R_k|\mathbf{R}_{-k}=1, \tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})}) \times \prod_{k=2}^{K} \text{OR}(R_k, \mathbf{R}_{\prec k}|\mathbf{R}_{\succ k}=1, \tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})})$$

其中，符号表示法和优势比与前文的定义一致。

我们有 $p\left(R_k|\mathbf{R}_{-k}=\mathbf{1}, \tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})}\right) = p\left(R_k|\mathbf{R}_{-k}=1, \tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})} \setminus \tilde{\mathbf{Z}}^{(\mathbf{r}_k=\mathbf{1})}\right)$。根据一致性，条件化竖线之后的每个反事实都等于观测代理变量。然后，成对OR

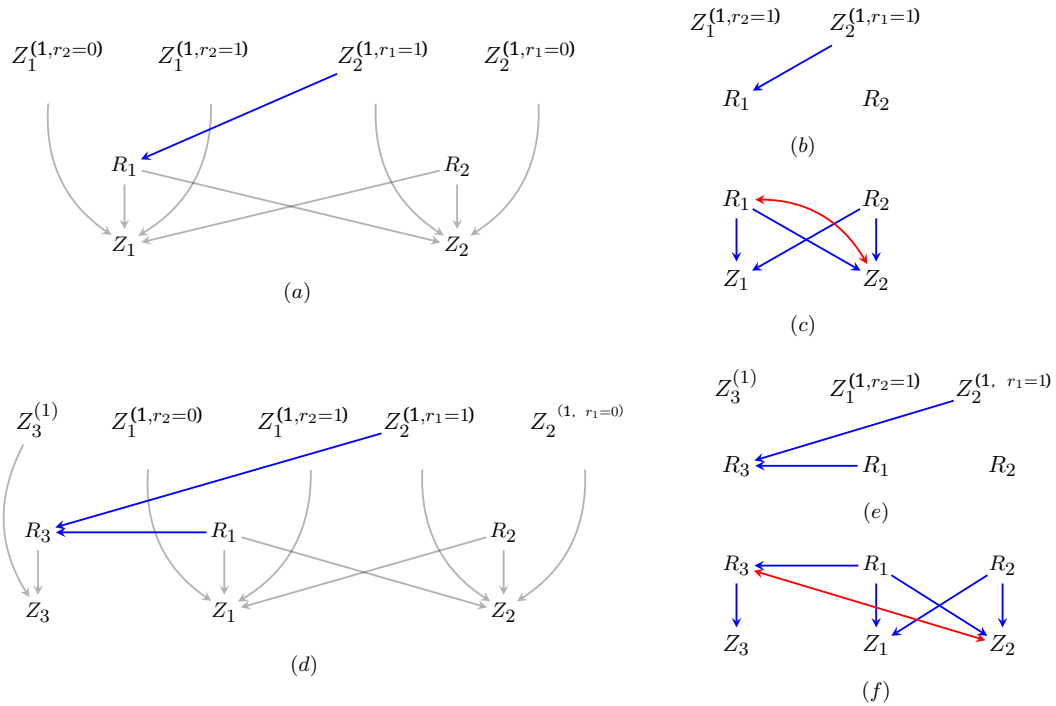


图11： P（$\tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})}$，$\mathbf{R}$）不可识别的示例。证明通过参数计数展示。(a)-(c)：扩展自我删失（邻居删失），(a)展示数据生成过程，(b)为 P（$\tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})}$，$\mathbf{R}$）的祖先有向混合图，(c)为观测分布的祖先有向混合图。(d)-(f)：扩展共谋者，(d)展示数据生成过程，(e)为 P（$\tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})}$，$\mathbf{R}$）的祖先有向混合图，(f)为观测分布的祖先有向混合图。

项已被识别，因为优势比（R_k，$R_i \mid \mathbf{R}_{-(k,i)} = 1$，$\tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})}$）不取决于 $\tilde{\mathbf{Z}}^{(\mathbf{r}_i=\mathbf{1})}$ 或 $\tilde{\mathbf{Z}}^{(\mathbf{r}_k=\mathbf{1})}$。最后，高阶项也以类似方式得到识别 [12, 89]。

完备性：证明完备性的一种方法是，统计对缺失数据图的完全可观测分布及其对应观测分布进行参数化所需的参数数量，并据此推断：如果前者所需参数多于后者，就不可能唯一识别完全可观测分布的所有参数。为此，我们假设所有变量均为二元变量，并采用嵌套马尔可夫模型的二元参数化。

首先，我们给出两个简单的 e-共谋路径示例，以展示参数计数论证的进行方式。随后，我们为所有图提供完备性的一般论证。

考虑图11(a)中的简单图，其中存在关于 $Z^{(1,r_1=1)}{}_2 \to R_1$ 的 e-自我审查。为了计算完整可观测性分布所需的参数数量，我们按照 [12] 中的步骤进行。缺失数据 ADMG完整分布的二元参数化，与普通无向化的有向无环图中的计数方式相同，即投影掉所有无关的反事实；但可以忽略确定性因子 $P(Z_i|R_j, Z^{(1,r_{\text{aff}(i)})}{}_i)$。由此得到图11(b)。与该图对应的5个参数为 $q(Z^{(1,r_2=1)}{}_1 = 0)$、$q(Z^{(1,r_1=1)}{}_2 = 0)$、$q(R_2 = 0)$ 和 $q(R_1 = 0|Z^{(1,r_1=1)}{}_2 = 1)$、$q(R_1 = 0|Z^{(1,r_1=1)}{}_2 = 0)$。

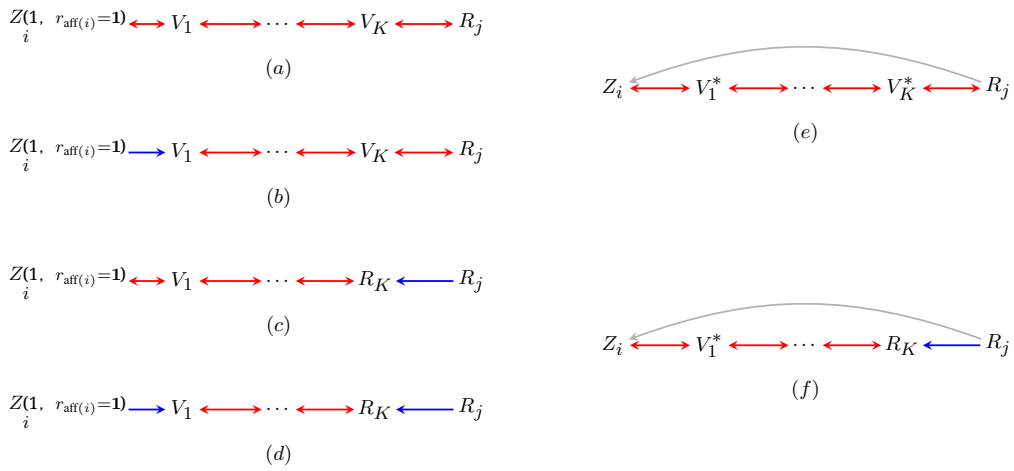


图 12：（a） $Z_i^{(1,r_{\text{aff}(i)}=\mathbf{1})}$ 与 R_j 之间的 e-共谋路径，其中 $R_j \in \mathbf{R}^{\text{aff}(i)}$ (i 和 j 是邻居）（b）投影掉 $Z_i^{(1,r_{\text{aff}(i)}=\mathbf{1})}$

接下来，通过投影掉 (a) 中的所有反事实，我们得到表示观测分布的图 (c)。在纠缠缺失性情形下计算完全可观测性下的观测分布时，需要特别注意：如果 Z_i 出现在某个参数的头部中，且相应的任一缺失指示变量 $R_{\text{aff}(i)}$ 出现在尾部中，则必须将核限制在以下情形：$R_i = 1$ 且 $\mathbf{R}_{\text{aff}(i)} = \mathbf{1}$。这是因为：(1) 当 $R_i = 0$时，头部取除 $Z_i =$ 之外的任何值的概率都确定为 0；(2) 将 $\mathbf{R}^{\text{aff}(i)}$ 设为不同于 1 的值的情形，与识别完整可观测性分布无关[11]。图 (c) 中的观测分布只需用4个参数即可参数化，分别对应于 $q(R_1 = 0)$、$q(R_2 = 0)$、$q(R_1 = 0, Z_2 = 0|R_2 = 1)$ 和 $q(Z_1 = 0|R_1 = 1, R_2 = 1)$。这比完整可观测性分布少一个参数，因此后者无法由数据唯一确定。

其次，考虑图 11(d) 中的图，其中有一个 e-共谋者 $Z^{(1,r_1=1)}{}_2 \rightarrow R_3 \leftarrow R_1$。完全可观测性分布如 (e) 所示，观测分布如 (f) 所示。对 (e) 进行计数可得参数总数为 9：$q(Z_3^{(1)} = 0)$、$q(Z^{(1,r_2=1)}{}_1 = 0)$、$q(Z^{(1,r_1=1)}{}_2 = 0)$、$q(R_1 = 0)$、$q(R_2 = 0)$，最后是 $q(R_3 = 0|R_1, Z^{(1,r_1=1)}\ 2)$，后者包含 4 个参数。另一方面，观测分布只需要 8 个参数：$q(R_1 = 0)$、$q(R_2 = 0)$、$q(R_3 = 0|R_1 = 1)$、$q(R_3 = 0|R_1 = 0)$、$q(Z_3 = 0|R_3 = 1)$、$q(Z_1 = 0|R_1 = 1, R_2 = 1)$、$q(Z_2 = 0|R_1 = 1, R_2 = 1)$ 和$q(Z_2 = 0, R_3 = 0|R_1 = 1, R_2 = 1)$。因此，无法根据观测数据唯一地映射回完整可观测性分布。最后，我们给出任意图的一般论证。

˜假设在$\mathbf{Z}$中有 n 个变量。为简单起见，假设所有反事实彼此独立，即 $Z_i^{(1,\mathbf{r}_{\text{aff}(i)}=\mathbf{r})} \perp\!\!\!\perp Z_i^{(1,\mathbf{r}_{\text{aff}(i)}=\mathbf{r}')}$，当$\mathbf{r} \neq \mathbf{r}'$ 且 $Z^{(1,\mathbf{r}_{\text{aff}(i)}=\mathbf{r})}\ i \perp\!\!\!\perp Z^{(1,\mathbf{r}_{\text{aff}(i)}=\mathbf{r}')}\ j$ 时；当 $\mathbf{r}$ 可能等于也可能不等于$\mathbf{r}'$时。图中，以下对象之间没有边

[11]所需考虑的是，$R_i = 1$ 始终成立，但$\mathbf{R}\text{aff}\ (i)\ = \mathbf{1}$ 仅用于识别完整可观测性分布。对于其他情况，所需考虑的因素各不相同。

反事实 [12]。根据假设1， $\tilde{\mathbf{Z}}^{(\mathbf{r}\neq\mathbf{1})}$ 不是任何 $R\in\mathbf{R}$的父节点，而且可以将其投影出去（不会添加新边），而不影响$\tilde{\mathbf{Z}}^{(\mathbf{r}=\mathbf{1})}$、$\mathbf{R}$ 或代理变量$\mathbf{Z}$中任何变量的马尔可夫毯。

图12（a）–（d）展示了 $Z^{(1,\mathbf{r}_{\text{aff}(i)}=\mathbf{1})}{}_i$ 与 R_j之间所有可能的 e-共谋路径。这里，$R_j\in\mathbf{R}^{\text{aff}(i)}$。假设有 K 个变量 $V_1\cdots V_K$位于 $Z^{(1,\mathbf{r}_{\text{aff}(i)}=\mathbf{1})}{}_i$ 与 R_j之间最短的碰撞路径上。要使其成为最短碰撞路径，我们要求当 i 与 l 具有相同的邻域时，即当$R_l\in\mathbf{R}^{\text{aff}(i)}$ 且 $R_i\in\mathbf{R}^{\text{aff}(l)}$时，$Z^{(1,\mathbf{r}_{\text{aff}(l)}=\mathbf{1})}{}_l$ 或 R_l 不在 $V_1,\cdots V_K$中。否则，我们可以截短该路径，使其成为连接 $Z^{(1,\mathbf{r}_{\text{aff}(i)}=\mathbf{1})}{}_i$ 与 R_j 的最短路径，并使 i 与 j 互为邻居。图12（e）展示了（a）和（b）的投影，（f）展示了（c）和（d）的投影。$V^*\in\mathbf{Z}\setminus Z_l$、$\mathbf{R}\setminus R_l$、$i$和 l 互为邻居。

现在，我们分别考察这些路径 (a)–(d) 及其对应的潜在投影 (e) 和 (f)，并假设它们出现在一个其余部分完全断开的更大图中。与前面处理示例时类似，我们统计莫比乌斯参数的数量（作为 K 的函数），并证明完整可观测性分布的参数总是比观测分布多。当我们将这些共谋路径置于一个具有任意连通性的更大图中时，由于仅 g-合谋路径所导致的差异，完整可观测性分布仍无法识别。也就是说，任何边超图（超模型）也都无法识别。

以下事实将用于统计二元模型中的参数数量：给定一条长度为 K' 的双向链 $V_1\leftrightarrow\cdots\leftrightarrow V_{K'}$，对该链进行参数化所需的参数数量为 $K'(K'+1)/2$，对应的参数如下：

$$
\begin{aligned}
&q(V_1=0)\ q(V_1=V_2=0)\ \cdots q(V_1=\cdots V_{K'}=0):\ K'\text{参数}\\
&q(V_2=0)\ q(V_2=V_3=0)\ \cdots q(V_2=\cdots V_{K'}=0):\ K'-1\text{参数}\\
&\qquad\qquad\qquad\qquad\qquad\qquad\cdots\\
&\qquad\qquad\qquad\qquad\qquad\qquad q(V_{K'}=0):1\text{参数}
\end{aligned}
$$

图 12 的参数计数12(a)、(b)、(e)

1. 图 1212(a) 中的莫比乌斯参数数量为 $(K+2)(K+3)/2$

   - 这是一个长度为 $K'=K+2$ 的双向链 $Z_{(1,\ \mathbf{r}_{\text{aff}(i)}=\mathbf{1})\,i}\leftrightarrow\cdots\leftrightarrow R_j$。

2. 图中莫比乌斯参数的数量12(b) 为 $(K+2)(K+3)/2$

   - $q(Z_{(1,\mathbf{r}_{\text{aff}(i)}=\mathbf{1})\,i}=0)$ 即 1 个参数。
   - 长度为 $K'=K$ 的双向链 $V_2\leftrightarrow\cdots\leftrightarrow V_K\leftrightarrow R_j$，即 $K(K+1)/2$个参数。
   - 涉及 V_1 的内在集，即 $q(V_1=0|Z_{(1,\mathbf{r}_{\text{aff}(i)}=\mathbf{1})\,i})$，$q(V_1=V_2=0|Z_{(1,\mathbf{r}_{\text{aff}(i)}=\mathbf{1})\,i})$，$\cdots$，$q(V_1=V_2=\cdots R_j=0|Z_{(1,\mathbf{r}_{\text{aff}(i)}=\mathbf{1})\,i})$ 每个包含 2 个参数，因此总计为$2*(K+1)$ 个参数。

3. 图 12(e) 中莫比乌斯参数的数量为 $(K+2)(K+3)/2-1$

[12]这一假设不会影响其他类型模型的完备性一般性，因为该模型是其他模型的子模型，而这些模型涉及反事实之间的依赖；在此模型中，完备性保证了其他模型中的完备性。

- 长度为 $K'=K+1$ 的双向链 $V_1^* \leftrightarrow \cdots \leftrightarrow R_j$ 有 $(K+1)(K+2)/2$ 个参数。需要注意的是，对于任意 $Z_l \in \{V_1 \cdots V_K\}$，在莫比乌斯参数的头部中，如果$R \in R_l \cup \mathbf{R}_{\text{aff}(l)}$是该参数的尾部，$R$ 总是确定性地设为 1。因此，这归结为计算简单双向链。

- 涉及 Z_i 的内在集数量为 $K+1$（而非 $K+2$），因为 R_i 不可固定，且集合 $\{Z_i, V_1^*, \cdots, V_K^*\}$ 不是内在集。每个内在集对应两个参数，因此共有 $2*(K+1)$ 个参数。

图 12(c)、(d)、(f) 的参数计数

1. 图 12(c) 中的莫比乌斯参数数量为 $(K+2)(K+3)/2$
   - $q(R_j=0)$，即 1 个参数。
   - 长度为 $K'=K$ 的双向链 $Z_{(1,\ \mathbf{r}_{\text{aff}(i)}=\mathbf{1})_i} \leftrightarrow \cdots V_{K-1}$，因此有 $K/(K+1)/2$ 个参数。
   - 涉及 R_K 的内在集，即 $q(R_K=0|R_j)$、$q(R_K=V_{K-1}=0|R_j)$、$\cdots$, $q(R_K=V_{K-1}=\cdots, Z_{(1,\mathbf{r}_{\text{aff}(i)}=\mathbf{1})_i}=0|R_j)$，也就是 $2*(K+1)$ 个参数。
2. 图中莫比乌斯参数的数量12(d) 为 $(K+2)(K+3)/2$
   - $q(Z_{(1,\mathbf{r}_{\text{aff}(i)}=\mathbf{1})_i}=0)$ 即 1 个参数。
   - $q(R_j=0)$，即 1 个参数。
   - 长度为 $K'=K-2$ 的双向链 $V_2 \leftrightarrow \cdots \leftrightarrow V_{K-1}$，即 $(K-2)(K-1)/2$ 个参数。
   - 涉及 V_1 且不涉及 R_K 的内在集，即 $q(V_1=0|Z_{(1,\text{raff}(i)=\mathbf{1})_i})$、$q(V_1=V_2=0|Z_{(1,\text{raff}(i)=\mathbf{1})_i})$、$\cdots$, $q(V_1=V_2=\cdots V_{K-1}=0|Z_{(1,\text{raff}(i)=\mathbf{1})_i})$，每个对应 2 个参数，因此共计 $2*(K-1)$ 个参数。
   - 涉及 R_K 而不涉及 V_1的内在集，即 $q(R_K=0|Z_{(1,\text{raff}(i)=\mathbf{1})_i})$、$q(R_K=V_2=0|Z_{(1,\text{raff}(i)=\mathbf{1})_i})$、$\cdots$, $q(R_K=V_2=\cdots V_{K-1}=0|Z_{(1,\text{raff}(i)=\mathbf{1})_i})$，每个各对应 2 个参数，因此总计 $2*(K-1)$ 个参数。
   - 同时涉及 V_1和 R_K的唯一一个内在集，即 $q(V_1=V_2=,\cdots=R_K=0|R_j, Z_{(1,\mathbf{r}_{\text{aff}(i)}=\mathbf{1})_i})$ 对应 4 个参数。
3. 图 12(f) 中的莫比乌斯参数个数为 $(K+2)(K+3)/2-1$
   - $q(R_j=0)$，即 1 个参数。
   - 长度为 $K'=K$ 的双向链 $Z_i \leftrightarrow V_1^* \leftrightarrow \cdots \leftrightarrow V_{K-1}$ 有 $(K)(K+1)/2$ 个参数。和之前一样，对于任意 $Z_l \in \{V_1 \cdots V_{K-1}\}$，在莫比乌斯参数的头部中，如果$R \in R_l \cup \mathbf{R}_{\text{aff}(l)}$ 是该参数的尾部，则始终将 R 确定性地设为 1。
   - 涉及 R_K 的内在集，即 $q(R_K=0|R_j)$,$q(R_K, V_{K-1}=0|R_j)$，$\cdots$, $q(R_K, V_{K-1}, \cdots, V_1=0|R_j)$，对应 $2*K$ 个参数；以及内在集 $q(R_K, V_{K-1}, \cdots, V_1=0, Z_i=0|R_j=1)$，它只对应 1 个参数（而非 2 个），因为 $R_j \in$Raff(i)，因此必须设为 1。

□

参考文献

[1] Donald B. Rubin。推断与缺失数据。*Biometrika*，63(3):581–592，1976年。[2] Roderick J Little。缺失数据假设。《统计学及其应用年度评论》，8:89–107，2021年。[3] R.J.A. Little 和 D.B. Rubin。《缺失数据的统计分析》。Wiley 概率与统计丛书。Wiley，2002年。[4] M Maria Glymour。运用因果图理解社会流行病学中的常见问题。《社会流行病学方法》，第393–428页，2006年。[5] Rhian M. Daniel、Michael G. Kenward、Simon N. Cousens 和 Bianca L. De Stavola。运用因果图指导缺失数据问题的分析。*Statistical Methods in Medical Research*，21(3):243–256，2012年。[6] Fernando Martel García。随机对照实验中问题性脱落的定义与诊断。可见于 *SSRN 2302735*，2013年。[7] Karthika Mohan、Judea Pearl 和 Jin Tian。用于缺失数据推断的图模型。《神经信息处理系统进展》，26，2013年。[8] Felix Thoemmes 和 Norman Rose。关于可能增加缺失数据问题偏差的辅助变量的审慎说明。《多元行为研究》，49(5):443–459，2014年。[9] Jin Tian。图模型中的随机缺失。载于人工智能与统计学，第977–985页。PMLR，2015年。[10] Ilya Shpitser。对非单调且非随机缺失数据函数的一致估计。《神经信息处理系统进展》，29:3144–3152，2016年。[11] Rohit Bhattacharya、Razieh Nabi、Ilya Shpitser 和 James Robins。有向无环图表示的缺失数据模型中的识别。载于《人工智能不确定性会议第三十五届会议论文集（第*35*届不确定性人工智能会议）》。AUAI Press，2019年。[12] Razieh Nabi、Rohit Bhattacharya 和 Ilya Shpitser。缺失数据图模型中的完整分布识别：完备性结果。载于国际机器学习会议，第7153–7163页。PMLR，2020年。[13] Karthika Mohan 和 Judea Pearl。用于处理缺失数据的图模型。《美国统计学会杂志》，第1–16页，2021年。[14] Daniel O Scharfstein、JaronJR Lee、Aidan McDermott、Aimee Campbell、Edward Nunes、Abigail G Matthews 和 Ilya Shpitser。对非单调缺失二元结局的随机试验进行马尔可夫约束分析：敏感性分析与识别结果。*arXiv* 预印本 *arXiv:2105.08868*，2021年。[15] Razieh Nabi 和 Rohit Bhattacharya。关于缺失数据模型的可检验性与拟合优度检验。*arXiv* 预印本 *arXiv:2203.00132*，2022年。[16] Razieh Nabi、Rohit Bhattacharya、Ilya Shpitser 和 James Robins。缺失数据模型的因果与反事实视角。*arXiv* 预印本 *arXiv:2210.05558*，2022年。[17] P. M. Aronow 和 C. Samii。在一般干扰下估计平均因果效应。技术报告，2013年。[18] S. Athey、D. Eckles 和 G.W. Imbens。网络干扰下的精确 p值。《美国统计学会杂志》，113 (521):230–240，2018年。

[19] G. Basse、A. Feller 和 P. Toulis。存在干扰时因果效应的随机化检验。*Biometrika*，106 (2):487–494，2019年。 [20] G. W. Basse 和 E. M. Airoldi。存在网络相关结果时的模型辅助实验设计。*Biometrika*，105 (4):849–858，2018年。 [21] J. Bowers、F. M. M 和 P. C. 关于个体间干扰的推理：通用框架。《*Political Analysis*》，21:97–124，2013年。 [22] X. Cai、W. W. Loh 和 F. W. Crawford。传染情况下的因果干预效应识别。 *arXiv* 预印本 *arXiv:1912.04151*，2019年。 [23] Daniel J Eck、Olga Morozova 和 Forrest W Crawford。传染病干预易感性效应的随机化。《数学生物学杂志》，85(4):1–22，2022年。 [24] Dean Eckles、Brian Karrer 和 Johan Ugander。网络实验的设计与分析：减少干扰造成的偏差。《因果推断杂志》，5(1)，2017年。 [25] Laura Forastiere、Edoardo M Airoldi 和 Fabrizia Mealli。网络观察性研究中处理效应与干扰效应的识别和估计。《美国统计学会杂志》，116(534):901–918，2021年。 [26] Razieh Nabi、Joel Pfeiffer、Denis Charles 和 Emre Kıcıman。赞助搜索广告中存在干扰时的因果推断。《大数据前沿》，5，2022年。 [27] Bryan S. Graham、Guido W. Imbens 和 Geert Ridder。社会溢出效应存在时隔离效应的测量：一种非参数方法。技术报告，美国国家经济研究局，2010年。 [28] M. Elizabeth Halloran 和 Claudio J. Struchiner。传染病中的因果推断。*Epidemiology*，第142–151页，1995年。 [29] M Elizabeth Halloran 和 Michael G Hudgens。疫苗对传染性的效应的因果推断。《国际生物统计学杂志》，8(2):1–40，2012年。 [30] Guanglei Hong 和 Stephen W. Raudenbush。幼儿园留级政策评估：多层观察数据因果推断案例研究。《美国统计学会杂志》，101(475):901–910，2006年。 [31] Michael G Hudgens 和 M Elizabeth Halloran。迈向存在干扰时的因果推断。《美国统计学会杂志》，103(482):832–842，2008年。 [32] Ravi Jagadeesan、Natesh S Pillai 和 Alexander Volfovsky。存在干扰的网络中估计处理效应的设计。《统计年鉴》，48(2):679–712，2020年。 [33] Panos Toulis 和 Edward Kao。因果同伴影响效应的估计。载于国际机器学习会议，第 1489–1497 页。PMLR，2013年。 [34] MichaelP Leung。网络干扰下的处理效应与溢出效应。《经济学与统计学评论》，102(2):368–380，2020年。 [35] Lan Liu 和 Michael G Hudgens。存在干扰时因果效应的大样本随机化推断。《美国统计学会杂志》，109(505):288–301，2014年。 [36] Georgia Papadogeorgou、Fabrizia Mealli 和 Corwin M Zigler。针对集群和总体层面的处理分配方案，对存在干扰的个体进行因果推断。《生物统计学》，75(3):778–787，2019年。 [37] David Puelz、Guillaume Basse、Avi Feller 和 Panos Toulis。一般干扰下因果效应随机化检验的图论方法。 *arXiv* 预印本 *arXiv:1910.10862*，2019年。 [38] Paul R. Rosenbaum。随机实验中个体间的干扰。《美国统计学会杂志》，102(477):191–200，2007年。

[39] Donald B Rubin. 评论：Neyman（1923）与实验和观察性研究中的因果推断。*Statistical Science*，5(4):472–480，1990年。[40] Fredrik Sävje。设定错误的暴露映射下的因果推断。*arXiv*预印本 *arXiv:2103.06471*，2021年。[41] Fredrik Sävje、Peter Aronow 和 Michael Hudgens。未知干扰下的平均处理效应。*Annals of Statistics*，49(2):673，2021年。[42] Michael E Sobel。住房流动的随机化研究揭示了什么？面对干扰的因果推断。《美国统计学会杂志》，101(476):1398–1407，2006年。[43] Eric J Tchetgen Tchetgen 和 Tyler J VanderWeele。存在干扰时的因果推断。*Statistical Methods in Medical Research*，21(1):55–75，2012年。[44] Panos Toulis、Alexander Volfovsky 和 Edoardo M Airoldi。处理之间存在网络纠缠时的倾向评分方法。*arXiv* 预印本 *arXiv:1801.07310*，2018年。[45] Tyler J VanderWeele。基于邻域的聚类与纵向数据的直接效应和间接效应。*Sociological Methods & Research*，38(4):515–544，2010年。[46] Changgee Chang、Yi Deng、Xiaoqian Jiang 和 Qi Long。分布式健康数据网络中不完整数据分析的多重插补。*Nature Communications*，11(1):1–11，2020年。[47] Jeffrey A Smith、James Moody 和 Jonathan H Morgan。网络抽样覆盖率 II：非随机缺失数据对网络测量值的影响。*Social Networks*，48:78–99，2017年。[48] Krista J Gile 和 Mark S Handcock。缺失数据网络分析及其在全国青少年健康纵向研究中的应用。*Journal of the Royal Statistical Society: Series C(Applied Statistics)*，66(3):501–519，2017年。[49] Zack W Almquist 和 Carter T Butts。缺失数据下的动态网络分析：理论与方法。*Statistica Sinica*，28(3):1245–1264，2018年。[50] Bruce Sacerdote。教育中的同伴效应：它们可能如何发挥作用、影响有多大，以及迄今为止我们了解多少？载于*Handbook of the Economics of Education*，第3卷，第249–277页。Elsevier，2011年。[51] Judea Pearl。《因果关系：模型、推理与推断》。剑桥大学出版社，第2版，2009年。ISBN 978-0521895606。[52] Marloes Maathuis、Mathias Drton、Steffen Lauritzen 和 Martin Wainwright。《图模型手册》。CRC出版社，2018年。[53] Thomas S. Richardson 和 Jamie M. Robins。单世界干预图（SWIG）：对因果关系的反事实方法与图形方法的统一。预印本，2013年。[54] Judea Pearl。《智能系统中的概率推理》。Morgan and Kaufmann，圣马特奥，1988年。[55] Judea Pearl。《因果关系》。剑桥大学出版社，2009年。[56] Miguel A Hernan 和 James M Robins。《因果推断：假如》。CRC出版社：Taylor and Francis Group，2020年。[57] Thomas S. Verma 和 Judea Pearl。因果模型的等价性与综合。技术报告 R-150，计算机科学系，加利福尼亚大学洛杉矶分校，1990年。

[58] Thomas S. Richardson。无环有向混合图的马尔可夫性质。*Scandinavial Journal of Statistics*，30(1):145–157，2003年。[59] Jin Tian 和 Judea Pearl。关于含隐变量因果模型的可检验推论。载于*Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence (UAI-02)*，第18卷，第519–527页。AUAI Press，俄勒冈州科瓦利斯，2002年。[60] Ilya Shpitser 和 Judea Pearl。递归半马尔可夫因果模型中联合干预分布的识别。载于*Proceedingsof the Twenty-First National Conference on Artificial Intelligence (AAAI-06)*。AAAI Press，帕洛阿尔托，2006年。[61] Yimin Huang 和 Marco Valtorta。Pearl 干预演算是完备的。载于*Twenty Second Conference On Uncertainty in Artificial Intelligence*，2006年。[62] Thomas S. Richardson、Robin J. Evans、James M. Robins 和 Ilya Shpitser。无环有向混合图的嵌套马尔可夫性质，2017年。工作论文。[63] Rohit Bhattacharya、Razieh Nabi 和 Ilya Shpitser。含隐变量图模型中因果效应的半参数推断。《机器学习研究杂志》，23:1–76，2022年。[64] Elizabeth L. Ogburn 和 Tyler J. VanderWeele。干扰的因果图。*Statistical Science*，29(4):559–578，2014年。[65] M.G. Hudgens 和 M.E. Halloran。迈向存在干扰时的因果推断。《美国统计学会杂志》，103(482):832–842，2008年。[66] Eric J. Tchetgen Tchetgen 和 Tyler J. VanderWeele。存在干扰时的因果推断。*Statistical MethodsinMedical Research*，21(1):55–75，2012年。[67] Eric J Tchetgen Tchetgen、Isabel R Fulcher 和 Ilya Shpitser。网络上因果效应的自动 g 计算。《美国统计学会杂志》，116(534):833–844，2021年。[68] A.P. Dempster、N.M. Laird 和 D.B. Rubin。通过 EM 算法从不完整数据中进行极大似然估计。《皇家统计学会杂志，*B*辑》，39:1–38，1977年。[69] Nicholas J Horton 和 Nan M Laird。含缺失协变量的广义线性模型的极大似然分析。*Statistical Methods in Medical Research*，8(1):37–50，1999年。[70] Donald B Rubin。多重插补概述。载于*Proceedings of the survey research methods section of the American statistical association*，第79–84页。Citeseer，美国新泽西州普林斯顿，1988年。[71] Joseph L Schafer。多重插补：入门介绍。*Statistical methods in medical research*，8(1):3–15，1999年。[72] James M. Robins、Andrea Rotnitzky 和 Lue P. Zhao。某些回归变量并非始终可观测时回归系数的估计。《美国统计学会杂志》，89:846–866，1994年。[73] Lingling Li、Changyu Shen、Xiaochun Li 和 James M Robins。关于缺失数据的加权方法。*Statistical methods in medical research*，22(1):14–30，2013年。[74] Eric J Tchetgen Tchetgen、Linbo Wang 和 BaoLuo Sun。非单调非忽略缺失数据的离散选择模型：识别与推断。*StatisticaSinica*，28(4):2069，2018年。[75] Margaret C Wu 和 R. J. Carroll。通过对删失过程建模，估计并比较存在信息性右删失时的变化。《生物统计学》，175-188，1988年。

[76] Sheng Wang、Jun Shao 和 Jae Kwang Kim。利用工具变量方法对不可忽略的无响应进行识别与估计。*Statistica Sinica*，第1097–1116页，2014年。[77] Wang Miao、Peng Ding 和 Zhi Geng。非忽略缺失数据下正态模型与正态混合模型的可识别性。《美国统计学会杂志》，111(516):1673–1683，2016年。[78] Wang Miao 和 EricJ Tchetgen Tchetgen。关于带影子变量的非随机缺失下双重稳健估计量的多种形式。*Biometrika*，103(2):475–482，2016年。[79] BaoLuo Sun、Lan Liu、Wang Miao、Kathleen Wirth、James Robins 和 Eric J Tchetgen Tchetgen。利用工具变量对非随机缺失数据进行半参数估计。*Statistica Sinica*，28(4):1965，2018年。[80] Matthew O Jackson、Zhongjian Lin 和 Ning Neil Yu。在估计处理效应时，倾向评分中对同伴影响进行调整。*Available at SSRN 3522256*，2020年。[81] C Hendricks Brown。防范纵向研究中的非随机缺失数据。《生物统计学》，第143–155页，1990年。[82] Karthika Mohan、Felix Thoemmes 和 Judea Pearl。不完整数据下的估计：线性情形。载于*Proceedings of the International Joint Conferences on Artificial Intelligence Organization*，2018年。[83] Werner Ebeling 和 Rainer Feistel。自组织与进化物理学。John Wiley & Sons，2011年。[84] Hua Yun Chen。用于测量关联的半参数优势比模型。*biometrics*，63:413–421，2007年。[85] Ryan M Andrews 和 Vanessa Didelez。关于因果中介分析中跨世界独立性假设的见解。*Epidemiology*，32(2):209–219，2020年。[86] Anastasios Tsiatis。半参数理论与缺失数据。Springer-Verlag New York，第一版，2006年。[87] Steffan L. Lauritzen。图模型。牛津，英国：Clarendon，1996年。[88] Robin J. Evans 和 Thomas S. Richardson。离散数据的马尔可夫无环有向混合图。*Annals of Statistics*，第1–30页，2014年。[89] Daniel Malinsky、Ilya Shpitser 和 Eric J Tchetgen Tchetgen。非单调非随机缺失数据的半参数推断：无自我删失模

l.

《美国统计学会杂志》，第1–9页，2021年。