arXiv:2304.07638v1 [cs.LO] 15 Apr 2023

弦图中的因果模型

Robin Lorenz*、Sean Tull†

Quantinuum，英国牛津博蒙特街17号

摘要

由Pearl及其合作者，以及Spirtes、Glymour和Scheines率先提出的 *framework of causal models*，为因果推理提供了一种原则性方法，如今已应用于众多科学领域。

在此，我们以 *string diagrams*的语言介绍因果模型的框架，并使用范畴论对其进行形式化解释。一类称为 *network diagrams*的弦图与 *directed acyclic graphs*一一对应。通过将此类图的组件解释为对称幺半范畴中的随机映射、函数或更一般的信道（具有“复制-丢弃”结构（*cd-category*）），即可给出一个 *causal model* 。这将模型表示为单一的数学对象，既可凭直觉理解和推理，又具有严格性。

本工作以此前的研究为基础，尤其是 Fong 和 Jacobs、Kissinger 和 Zanasi，以及 Fritz 和 Klingler 的工作，在 cd-范畴中给出了因果模型和函数因果模型的图示定义，分别推广了因果贝叶斯网络和结构因果模型。我们形式化了对模型进行 *interventions* 的最一般类型，其中包括但不限于由 do-干预描述的原子干预，并提出了 *open causal model*的自然概念，即一种“带输入”的因果模型。

为了将这些概念用于因果推理，我们还提出了一种基于 *normalisationbox*的条件化方法，使因果推断计算能够完全以图示方式进行。我们在这一框架中用它们来定义*counterfactuals* ，并以完全图示化的方式处理因果效应和反事实的可识别性问题。

以这种方式呈现因果模型，既有助于探讨因果推理中的基础性问题，也尤其能发挥阐释作用并带来 *pedagogical value*。事实上，本手稿旨在让不同领域的读者都能理解，包括因果模型实践者和应用范畴论研究者。为阐释核心思想，文中以图示语言讨论了因果模型文献中的许多例子。

总的来说，我们论证并展示了：依据因果模型框架， *causal reasoning* 最自然、直观的做法是 *diagrammatic reasoning*。

目录

1 引言 3 1.1 概述与浅显导览.6

2 范畴论框架 **10**

2.1 弦图与对称幺半范畴.10 2.2 cd-范畴.12 2.3 确定性过程与帽.15 2.4 归一化.17 2.5 条件化与条件独立性.18

*robin.lorenz@quantinuum.com

†sean.tull@quantinuum.com

3 因果模型 **22**
3.1 定义.22 3.2 作为图示的因果模型.23 3.3 函数因果模型.27 3.4 忠实性概念.30

4 干预
31 4.1 定义.31 4.2 各种干预类型.32

5 开放因果模型 **36**
5.1 定义.36 5.2 开放模型的组合.39 5.3 因果模型的变换.42 5.3.1 内化与外化.42 5.3.2 开放与干预.44

6 含潜变量的因果模型的处理 **45**
6.1 潜在投影与祖先有向混合图.46 6.2 祖先有向混合图与网络图.47

7 因果效应的识别 **48**
7.1 问题.48 7.2 图示地呈现 c-分量条件——JKZ 结果.49 7.3 图示地呈现一般情形——示例.51

8 反事实 **60**
8.1 反事实概念.61 8.1.1 引入性示例.61 8.1.2 一般定义.63 8.2 反事实的可识别性.66 8.2.1 问题陈述与示例.66 8.2.2 简化的一般方法——simplify-cf 算法.72 8.2.3 识别算法 id-cf.75 8.3 反事实的推广.80

9 展望 **82**

10 致谢 **84**

参考文献 **84**

A 测度论概率 90

B 归一化 91 B.1 归一化的唯一性.91 B.2 推导归一化.93

C 条件化与条件独立性 95
C.1 命题 15 的证明.96 C.2 引理 16 的证明.97 C.3 引理 18 的证明.97 C.4 定理 19 的证明.99 C.5 定理 40 的证明.100 C.6 引理 42 的证明.101

D 修剪干预 E 定理90的证明 F 因果模型的分裂节点 **102**
表示 **102**
103

1 引言

因果推理对于许多领域的科学实践至关重要。尽管因果性质的问题无处不在，因此科学需要有原则的因果推理，但要建立扎实且富有成效的形式化方法，仍经历了漫长的历程。除了技术和社会学方面的挑战之外，还存在哲学层面的难题，因果关系几乎所有认识论和形而上学层面都曾引发争议 [1, 2, 3]。暂且搁置其中一些哲学问题的答案，由 Judea Pearl 及其合作者 [4], 以及 Spirtes、Glymour 和 Scheines [5], 开创的因果模型框架，在澄清并形式化科学实践中的因果推理方面，构成了巨大的技术进步。

本文基于众多作者正在开展的一项持续项目 [6, 7, 8, 9, 10, 11, 12, 13], ，提出对因果模型框架的重新表述，旨在进一步提升概念上的清晰度、易懂性和直观性；为此，我们将因果模型和因果推理完全置于弦图的语言中，以图形方式重新表述。

因果模型。因果模型框架阐明了：给定一组变量 V，关于 V （其中一部分）的因果结构与经验数据如何相互制约。经验数据的性质可以是观察性的，也可以是干预性的；因果结构则由变量之间的一组直接因果关系构成，通常以有向图表示，最常见的是有向无环图（DAG）。

人们可以区分因果模型的两个重要且密切相关的变体。在假定底层具有决定论的情况下，结构因果模型（SCM）根据因果结构规定具有依赖关系的函数式因果机制。在更粗粒度的层面上，因果知识较为不详尽时，因果贝叶斯网络（CBN）这一派生概念规定随机因果机制；这些机制共同定义一个相对于该 DAG 满足因果马尔可夫条件的概率分布。这些规定随后成为以有原则的方式作出因果判断和回答查询的基础。

其中的成就包括因果发现算法：在给定观察数据和某些假设的情况下，尽可能推断因果结构。另一项成果体现在因果推断问题的解决方案上：在已知因果知识的情况下，人们可以询问何时以及如何预测在进行干预并且例如强制某些变量取特定值时会发生什么，同时还能够对反事实问题作出无歧义的回答。该框架描绘了一个精确的因果层级，包含三种层次的问题：根据所拥有的数据类型，可以回答的问题从仅涉及关联关系的问题，逐步提升到干预性推理，再到反事实推理 [4, 14]。

过去四十年来，因果模型框架已发展成为一个丰富且高度跨学科的研究领域，其范围从计算机科学、统计学和哲学等原始学科，扩展到具体科学领域和工业应用，涵盖医疗健康、经济学和政策制定等方面（例如，参见参考文献 [15, 16, 17, 18] ，仅列举少数）。近年来，也涌现出大量研究机器学习背景下因果性的著作（参见参考文献 [19] 及下文的进一步讨论），并且有人尝试从量子力学基础中的因果性研究中提炼出对（经典）因果推断有用的技术经验[20, 21]。

这一形式体系同时运用图论和概率论。因果模型通常涉及一个有向无环图 G 和一个概率分布 P，但从某种意义上说，两者彼此独立——研究者考察有向无环图的性质，这些性质可能使某些概率计算成为可能；反过来，也可以检查概率等式，并据此修改图。为了追踪这两个对象如何相互约束，这可能需要相当复杂的记号和记账方式，因此不禁让人思考，能否用一个数学实体来处理这两个方面。

图示推理的力量。过去二十年间，另一个看似毫不相关的成功故事，是范畴论和图示推理在数学中的应用，

以及在计算机科学、量子计算和量子理论基础中的应用。范畴论可被视为复合的数学语言，它强调实体之间的过程或关系（称为态射），而非实体本身 [22]。根据所采用的范畴不同，这些过程可以有多种类型，包括物理过程、逻辑过程和计算过程 [23]。

其中一个重要方面，是用幺半范畴来形式化科学主题。幺半范畴也称为过程理论，并具有与之相关的图形演算，使人们能够借助直观的弦图，严谨地推理过程 [24]。

图形化方法在量子力学基础 [25] 和量子计算 [26]，的应用中取得了巨大成功，并构成了 ZX演算 [27] 的基础。ZX演算用于描述量子电路，如今已在工业中得到广泛应用。近年来，应用范畴论 [28] 这一新兴领域不断发展，将组合式方法拓展到更多科学领域，包括概率论 [8, 10]，、机器学习 [29]，、博弈论 [30] 和人工智能 [31]。

先前研究。由此自然会产生一个问题：因果模型框架能否用这种图示语言来表达，并通过清晰的范畴论形式化加以阐明？过去十年间，多项研究提出了这一问题，并为解决它迈出了重要步伐。

Coecke 和 Spekkens 在文献 [6] 中率先以图示方式呈现贝叶斯推断，并利用了幺半范畴可能具有的一种特殊结构，称为伴随弗罗贝尼乌斯结构。与这项工作密切相关，Fong 随后提出了一种旨在形式化因果推理 [7] 的贝叶斯网络范畴论。该方法以有向无环图生成一个自由范畴，即其因果理论；特别是，它还生成一个弦图，其连线对应于父节点集，并可根据其所解释的范畴赋予不同的语义。

此后，若干作者，尤其是 Jacobs 等人 [32, 33, 8, 34, 8] 和 Fritz [10, 35]，发展了范畴概率论，借助 cd-范畴和马尔可夫范畴这一简单框架，阐明了概率推理在结构上的本质（无需完整的伴随弗罗贝尼乌斯结构）。在这一框架下，Fritz 和 Klingler [12, 13] 最近还提出了范畴 d-分离——一种图解概念，可将通常的 d-分离概念加以泛化——并证明，对于马尔可夫范畴中条件独立性的相应泛化，该概念是可靠的。

在这些进展的基础上，并基本沿用 Fong 在参考文献 [7]，中提出的弦图框架，Jacobs、Kissinger和 Zanasi 在参考文献 [9, 36] 中不再局限于贝叶斯推理，而是转向干预、因果效应和反事实等明确的因果概念，并用弦图加以表达。本文将对他们的工作进行更详细的讨论。

本工作。尽管这些努力确立了一种以图示方式处理因果推理的有前景的方法，但它们仍较为零散，且往往要么侧重于相当形式化的范畴论层面，要么只涵盖因果模型框架的一个子集。缺少的是一份全面且易于理解的介绍，以字符串图语言阐述因果模型的基本框架。本文提供了这样一种处理方式。可以将其视为对 Jacobs 等人在参考文献 [36]，中工作的扩展，但不坚持严格区分句法与语义。本文：

- 给出了因果贝叶斯网络和结构因果模型（及其推广）的图示定义；
- 以完全一般的方式处理干预，也就是说，讨论与因果模型定义相容的最一般类型的干预，而不局限于原子干预，例如软干预和条件干预 [37];
- 将贝叶斯条件化纳入图解演算，使得因果推理问题中的计算——这类问题通常包含某种条件化——可以完全通过图示完成；值得注意的是，这种处理考虑到了条件分布所依据的给定分布可能缺乏全支撑的情况——这里将条件分布视为部分信道——同时仍能得到一种良定义的条件独立性概念，满足半拟阵公理，并保证 d-分离的可靠性；
- 介绍了开放因果模型这一概念——它们自然地源自某一类干预（包括 do-干预），并且与参考文献 [38] 中的“条件有向无环图”模型以及

参考文献 [39, 40] 中的开放图有关——此外还介绍了（开放）因果模型的变换，从而得到（开放）因果模型范畴；

- 讨论如何从基于无环有向混合图的潜变量模型（该模型刻画了潜在共同原因）转化为弦图表示；
- 讨论因果效应可识别性，并对参考文献 [36] 中的工作作了适度扩展；
- 探讨反事实，尤其是以图示术语给出反事实的一般定义，具有阐释和教学价值；此外，还以图示方式探讨反事实的可识别性问题，包括 make-cg 和 IDC* 算法的类似版本 [41]。

我们强调，本文的目标不是追求过度形式化（更形式化的基础见参考文献 [12] ），而是让不同领域的读者都能理解——因此，部分内容会针对特定读者作入门介绍——并展示因果模型的弦图表示的实用价值。在更详细地介绍本文内容之前，我们先进一步阐述一些研究动机。

动机：用弦图术语对因果模型框架进行范畴论表述，出于多种原因具有重要价值。

(1) 直观表示带来的益处。图示方法的优势在于，它能让人优雅地将构成因果模型的所有数据——因果结构，以及与因果机制相关的随机或确定性映射——整合到一个数学对象中。重要的是，这种语言不仅能直观地表示模型，还能将人们通常对模型所做的一切形式化。这包括推理规则（重写规则），它们可直接应用于弦图，使许多计算变得相当直接，并避免原本繁复的记账。弦图似乎是进行因果推理最自然的语言。

既然许多因果推断问题都有算法解法，为什么还要在意优雅的表示——难道我们不能把一切都交给机器来做吗？首先，科学史告诉我们，找到恰当的语言有助于推动进步，并加深我们的概念理解。其次，对于希望用纸笔推演具体示例的因果模型实践者来说，这当然很重要。第三，这也具有教育意义。无论如何强调这一框架的重要性都不为过——因果模型将长久存在 [42] ——因此我们或许可以期待，未来因果推理的基础知识会成为任何科学教育的一部分，甚至在学校里教授。为此，我们希望使用尽可能简单直观的语言。这与近期借助图示方法提升人们对量子物理和量子计算的基本“素养”的努力相呼应，参见文献 [43]。

（2）因果性与机器学习。因果性文献中一个源自因果层级的重要启示是：基于纯粹关联数据的机器学习（ML）存在根本性局限 [44]。在很大程度上被机器学习界忽视之后，机器学习中的因果性研究在过去几年里获得了重要地位，并开辟了富有前景的新方向（参见 Sch¨olkopf 等人的研究[19, 45, 46] ，其中有综述）。这两个领域之间的互动大致有两个方面。

首先，有人认为，以一种结合现代机器学习自“统计转向”以来的成功经验与因果模型思想的方式学习因果机制，有助于解决当前机器学习面临的重大挑战，例如在有限资源条件下实现（强）泛化、鲁棒性和效率，以及可解释性与公平性 [19]（例如参见参考文献 [47, 48, 49]）。人们寄希望于这样一种模型的潜在益处：该模型能够学习结构化表示，其中的结构对应于世界中的（因果）结构：

> 因此，能够稳健地解决现实世界中一系列问题的未来人工智能模型，很可能需要重复使用组件，这要求它们在不同任务和环境中都具备稳健性 *[.....]*。实现这一点的一种优雅方式，是采用一种模块化结构，使其映射现实世界中相应的模块化特性。换言之，如果现实世界确实具有模块化特征，即其中的组件／机制能够在一系列环境、任务和情境中发挥作用，那么，让模型采用相应的模块将是明智之举。 [19]

其次，因果模型框架将与因果分析相关的变量集合 O 视为既定，并未说明应如何确定这些因果关系项。在参考文献中提出的因果表征学习[1] 概念 [46, 19] 探讨了这样一个问题：机器学习模型能否首先学习这些因果关系项，即从较低层级的数据（例如像素）中抽象出来。具体应如何实现这一点，以及实践中需要满足哪些必要条件，目前仍不清楚，但已有大量工作在探索这一思想 [55, 56, 57, 58, 59, 60,61, 62, 63, 64, 65]。

如今，范畴论是研究组合（模块化）结构以及结构对应关系（例如模型与世界之间的对应关系）所使用的数学语言。人们也日益关注以范畴论和组合性为重点的机器学习与人工智能研究[66, 67], ，尤其是因果性在其中所发挥的作用 [68, 69, 65]。因此，鉴于因果性与机器学习之间可以通过这两种方式相互促进，首先以范畴论的形式表述因果模型，并着重展现因果结构作为组合结构的一面，只会有所裨益。

（3）基础层面。范畴论表述也应有助于解决因果性领域中的若干基础问题。首先，它有助于探讨一些哲学问题，例如是谁或是什么将世界划分为因果关系项，或者如何定义因果机制和干预。暂且搁置这些问题，并假定领域知识通常会给出没有争议的答案，这种做法无疑推动了因果模型的发展；然而，尝试寻找答案十分重要，这不仅对因果哲学如此，也很可能能为理论提供依据，从而界定并评估因果表征学习的适用范围（参见参考文献 [68, 70]）。

其次，为因果模型确定恰当的范畴论框架，尤其意味着要找出最一般、最抽象的结构，以便在此基础上定义“本质上”的因果推理。在构建因果模型框架的进一步推广时，这一点至关重要。cd-范畴或马尔可夫范畴的概念便发挥着这样的作用——见下文及参考文献 [8, 10, 35]。

第三，Schmid、Selby 和 Spekkens 在参考文献 [71]，中发展了因果推断理论框架，旨在“[解开] 因果关系与推断交织而成的乱麻”，也就是说，既要对二者进行形式化刻画，也要弄清它们如何相互作用，从而产生我们所熟知的因果推理。鉴于他们的框架已经采用过程理论术语，本篇对因果模型的阐述应有助于理清因果模型与因果推断理论之间的关系——这一问题已在参考文献 [71] 中提出。

第四，尽管因果推断理论在经典情形下也颇有意义，但它最初是为了解开量子理论中因果关系与推断之间的纠缠而发展起来的。更广泛地说，如今规模庞大的“量子因果性”领域，致力于理解因果概念在多大程度上可应用于量子理论；量子理论的基础争议使这项任务尤为紧迫。参考文献 [72, 73, 74, 75, 76] 中提出的一个颇具前景且基础扎实的量子因果模型框架，明确关注经典因果模型的量子泛化，但目前还缺少同样完整且直观的图示呈现。造成这一缺失的原因既有技术层面的，似乎也有深层的概念层面因素 [77, 78]。因此，必须先充分发展经典情形，以便明确指出其中哪些方面可以量子泛化，哪些则不能。这一点对于反事实尤其重要；目前关于量子因果模型中的反事实研究甚少，而本文对（经典）反事实的图示呈现应能提供恰当的阐释与参考。

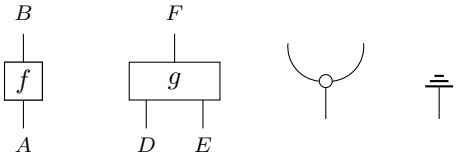
1.1 提纲与温和概述

现在让我们概述 论文中所介绍的因果模型的图解处理。

在全文中，我们的基本设定是带有某些额外结构的对称幺半范畴，这种结构使其成为所谓的 cd-范畴。我们强调，即使读者不熟悉范畴论的基础知识，也不必担心——对本文大多数内容而言，弦图的直观语言就已足够。范畴由对象以及对象之间的态射或过程构成，前者以导线表示，后者以方框表示。cd-范畴还带有复制映射，用于“共享”变量的值；也带有丢弃过程，用于表示边缘化，因此得名“cd”。以下展示了弦图中的几个组件（我们从下往上阅读）：一个从 A 到 B 的态射 f，另一个

[1]这项研究有其先驱，并且与 Chalupka 和 Eberhardt 对“因果抽象”和“因果粗化”的研究有关 *et al.*（例如，参考文献 [50, 51, 52, 53, 54]。

g，其输入为 D、E，输出为 F，以及一个复制映射和一个丢弃映射。



FStoch 是 cd-范畴的一个例子，其对象是有限集，态射是它们之间的随机映射；概率分布则是没有输入的特殊态射。我们将在第2节介绍范畴论框架。这部分内容是标准的，熟悉范畴概率论的读者可以略过，但第2.4节和第2.5节中关于概率条件化的定义或许值得一读。后者建立在先前的工作 [32, 33, 34, 8, 12, 13], 之上，不过我们的讲述专为本研究量身定制，并引入了一个归一化框，使得条件化可以完全以图示方式处理。

第3节给出了 cd-范畴 C 中因果模型的定义。普通的、针对离散变量的因果贝叶斯网络是其中一个实例，也是贯穿全文的主要例子。基本而言，给定一个集合 $\{X_i\}_i$，其中每个 X_i 都是 C 中的一个对象，其上的因果模型是一个有向无环图 G，该图的顶点为 $\{X_i\}_i$，并配有一组机制 $\{c_i\colon Pa(X_i) \to X_i\}_i$；其中，$Pa(X_i)$表示 X_i 在 G中的父节点集合。事实上，有向无环图可以等价地表示为一种特殊的字符串图，我们在此称之为网络图。粗略地说，这种图由机制对应的单输出盒子以及用于将变量输入其所有子节点的复制映射构成。于是，C 中的因果模型就是一个网络图，其中所有组成部分都在 C 中得到了解释。

作为一个基础的经典例子，考虑一个因果模型，它涉及以下变量：S 某人是否吸烟，L 他们是否会患上肺癌，B 一种可能是上述两者共同原因的遗传状况，以及 T他们肺部的焦油含量。变量 S、T、L 是观测到的，而 B未被观测到（潜在的）。图1a给出了一个合理的有向无环图，对应的因果模型则如图1b所示。

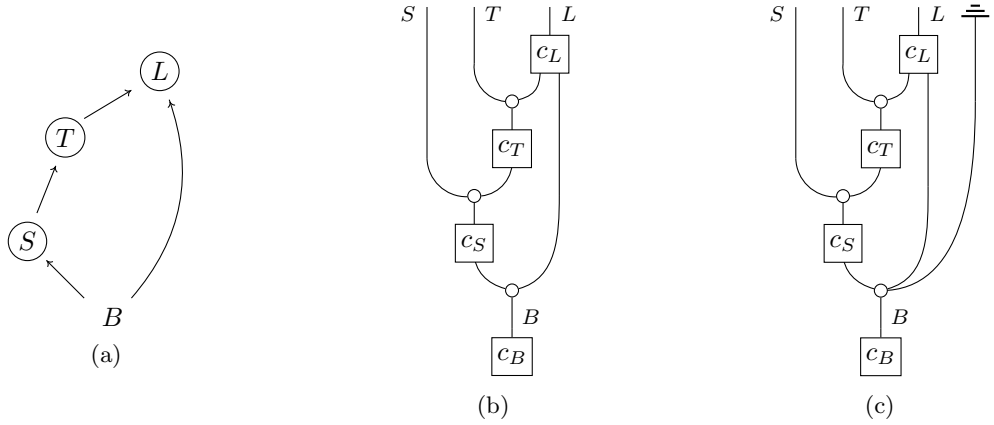


图1：有向无环图 G 见(a)，其中顶点外的环表示“观测到的”变量子集；(b)为对应的网络图，其中 $\circ$ 表示复制映射；(c)为等价图，其中 B 的导线通过丢弃它而被“边缘化”，以 表示。图示按自下而上读取。

网络图的输出线选择，尤其可以体现人们将哪些变量视为观测到的。图1c中的四个变量都作为输出，但随后 B 被丢弃。cd-范畴中的一条简单改写规则表明，图1b和图1c中的图示是等价的。事实上，图1b中的图示结构已经表明，若将其解释为 FStoch 中的对象，它表示一个关于 S、T、L、B的概率分布（的边缘分布），且该分布满足相对于 G的因果马尔可夫条件，即具有以下形式

$$P(S,T,L) \;=\; \sum_B P(L|T,B)\; P(T|S)\; P(S|B)\; P(B)$$

其中，$P(L|T,B)$、$P(T|S)$ 和 $P(S|B)$ 是随机映射 c_L、c_T 和 c_S 的密度（事实上，只要 P 具有全支撑，就可以将它们导出为 P 的贝叶斯条件分布）。

函数因果模型[2] 从形式上说是因果模型的一种特殊情况，并在第3.3节中讨论。这里，函数的概念被推广为 cd-范畴中的确定性态射。

因果模型的存在，而非仅仅是概率分布，尤其使我们能够回答关于对变量进行干预后所得分布将如何变化的问题。这里我们允许采用一种非常一般的干预：将模型的机制 $\{c_i\}_i$ 替换为定义在相同变量上的任意另一组机制 $\{c_i^*\}_i$，并且保持无环性。这与文献中对一般非原子干预的其他处理方式一致（例如，参见参考文献 [37]）。在第4节中，我们将讨论干预及其图示表现（参考文献 [36] 中称之为“图式手术”）。

实践中最常考虑的是所谓的 do-干预，即为一组变量赋予固定值。对上述示例执行 do-干预 $\mathrm{do}(S=s)$，相当于“断开” S 导线，丢弃其原有机制 c_S，并输入尖锐态 s（即在 FStoch 中，以 $S=s$ 为中心的点分布）。图 2 展示了结果 $P(S,T,L;\mathrm{do}(S=s))$。

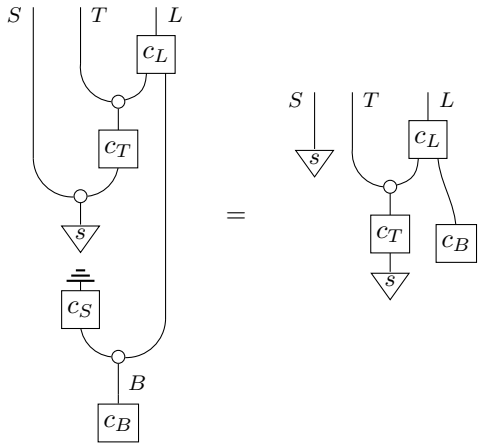
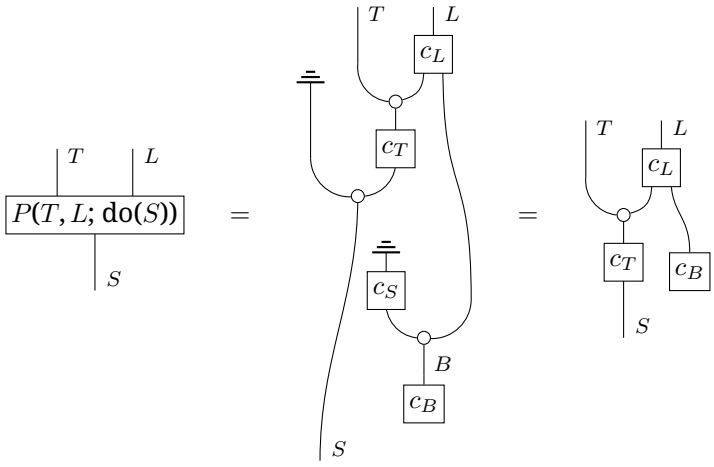


图2

如今，指定干预所真正需要的数据通常并非完整的因果模型，而只是其机制的一个子集。这自然引出了开放因果模型这一概念；它本质上是一个“带输入的模型”，我们将在第5节中介绍。一个例子是 $P(Y\,|\,\mathrm{do}(X))$，将其理解为从 x 映射到 $P(Y\,|\,\mathrm{do}(X=x))$ 的映射。对于上述例子，这使我们能够考虑如下内容：



(1)

开放模型的概念还使我们能够形式化因果模型的并行复合和顺序复合，以及比干预更一般的模型之间的变换。对此进行形式化

[2]也称为 *structural causal model* 和 *structural equational model*。

在第5.2节中，我们将证明开放因果模型本身构成一个对称幺半范畴。开放模型的构建基于与参考文献 [38] 中的条件有向无环图以及参考文献 [39, 40] 中的开放图相关的思想。

本文余下部分将转向这一框架的应用。在这些应用中，我们并不总是掌握因果结构的“完整”知识，有时只知道某些变量可能具有潜在共同原因。此类情形可用无环有向混合图（ADMG）表示；它是一种有向无环图（DAG），还可以包含表示潜在因果结构的双向边。第6节将讨论ADMG以及从DAG到ADMG的潜在投影概念。随后，我们介绍ADMG G的根化，这是一种有原则的方法，通过引入根节点将其映射到一个DAG，使得该DAG的潜在投影为 G 。图3中的ADMG与图1a中的DAG正是这种关系。根化ADMG的网络图为讨论可识别性问题提供了表述方式，而且重要的是，这样做不失一般性。[3]

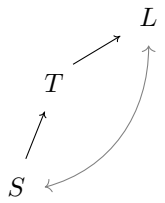


图3

可识别性问题的一个重要例子是因果效应的可识别性问题：我们何时以及如何才能仅凭观察数据和由ADMG指定的假定因果结构，唯一确定干预所产生的分布。第7节简要回顾这一问题，随后探讨如何用图示方式处理它。首先，我们重述Jacobs 等人发表于文献 [36] 中的一项结果；该结果以图示术语表达了Tian和Pearl [79], 提出的c-分量条件，至少适用于该条件成立的一种特殊情形。在这里，我们的设定显然可以将这一结果扩展到比原子干预更一般的干预。其次，我们讨论文献中的几个教学示例，展示用字符串图语言表示时，这些具体示例会变得多么简单。我们还将关注图示归一化框如何明确呈现这样一种相互作用：何时必须对趋零概率事件进行条件化，以及由此引发的可识别性问题。

因果模型框架的最后一个核心方面是反事实，即诸如“如果 Mary 服用了阿司匹林，她还会头痛吗？”这样的问题。在第8.1节中，我们以图示方式给出反事实的一般形式化定义，该定义允许反事实陈述中包含任意多个平行世界。尽管这与文献中的做法一致（例如，参见参考文献 [4, 41, 80]），我们认为，图示化的表述以及我们对定义反事实所需数据提出的精确条件具有澄清作用。尤其是，它清楚地阐明了反事实问题与“如果进行干预会怎样”这类问题之间的区别。

我们发现，以图示方式处理反事实既简单又自然，第8.2.1节中的一个示例对此作了展示。由此，我们在第8.2.2节提出了一个名为 simplify-cf 的算法，该算法通过应用重写规则来简化反事实图，其输出与参考文献 [41], 中的 make-cg 算法相对应，但可以说更加直接明了。随后，第8.2.3节介绍了一个名为 id-cf 的算法，该算法根据 simplify-cf 的输出评估反事实的可识别性，并以图示形式输出相应的识别表达式。这可以视为将参考文献 [41] 中算法解法的主要思想转换为图示形式。我们论证了该算法的可靠性，并讨论其完备性。

最后，第8.3节讨论了反事实概念的一种泛化，这种泛化在本文的设定中自然而然地出现：世界可以通过更一般的因果结构变化来定义，而不局限于 do-干预；此外，还可以对模糊事实进行条件化。

第9节以一系列有望推动未来工作的研究方向作结。

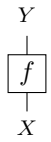
[3]至少对于许多问题而言，这样做不会损失一般性；不过，某些可识别性问题仍有待解决。但这些未决问题并非我们所采用的处理方式特有，而是更普遍地存在于因果模型文献中对可识别性问题的处理之中。详细讨论见后文各节。

2 范畴论框架

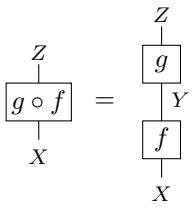
我们将首先介绍由众多作者发展起来的概率论范畴论方法，参见参考文献 [6, 8, 81]。形式上，这需要在“对称幺半范畴”或更具体地说“cd-范畴”中开展工作。实际上，这相当于使用直观而形式化的图示，即弦图 [24], ，来描述概率性过程以及（此处所说的因果）过程。对于本文而言，只需考虑由 $\mathbb{R}^+$ 值有限矩阵构成的范畴 $\mathsf{Mat}_{\mathbb{R}^+}$ ，如下文例 1 所述；不过，范畴论方法的适用范围要广泛得多。

2.1 字符串图与对称幺半范畴

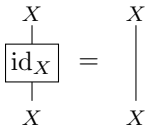
首先，我们从对称幺半范畴的基本概念，以及用于表示它们的字符串图语言开始。回顾一下，范畴 C 由一组对象 $X, Y, \ldots$ 和态射 $f\colon X \to Y$ 组成；这些态射彼此之间可以顺序复合。这样的态射也称为从 X 到 Y 的过程 f 。对象 X 表示为一条标记为 X的导线，而态射$f\colon X \to Y$则表示为一个方框，其输入 X 位于下方，输出 Y位于上方，图示由下向上阅读。



给定另一个态射 $g\colon Y \to Z$，我们可以构成它们的顺序复合 $g \circ f\colon X \to Z$，如下图所示。



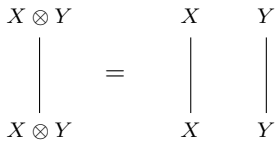
请注意，只有当态射的类型以这种方式匹配时，我们才能对它们进行顺序复合。范畴中的每个对象 X也都带有一个恒等态射 $\mathrm{id}_X\colon X \to X$ ，它表示为一条空白导线:



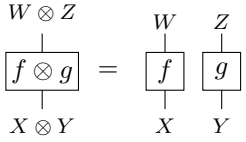
它充当复合的单位元 f 或复合：对任意 $f\colon X \to Y$，$\mathrm{id}_Y \circ f = f = f \circ \mathrm{id}_X$。注意，这条规则在图示语言中显而易见的。 age.

形式上，对称幺半范畴 $(\mathsf{C}, \otimes, I)$ 是一个范畴 C，它带有一个函子 $\otimes\colon \mathsf{C} \times \mathsf{C} \to \mathsf{C}$、一个指定对象 I ，以及若干自然变换，用以表达 $\otimes$ 满足适当的结合律和对称性，并且 I 是复合的单位元 [82]。这些特性用弦图表达最为自然，如下所示。

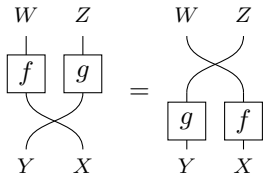
首先，对于任意一对对象 X, Y，我们可以构造它们的并行复合（或称“张量”） $X \otimes Y$，图中将它们的导线并排放置。



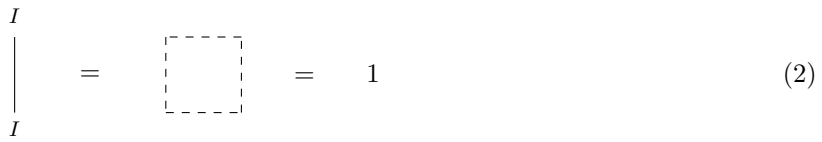
类似地，给定态射 $f\colon X \to W$和 $g\colon Y \to Z$，我们可以构造它们的并行复合 $f \otimes g\colon X \otimes Y \to W \otimes Z$，如下图所示。



张量具有对称性，这意味着我们也可以让成对的导线彼此交换，使得交换两次后导线保持不变；方框也可以沿着交换移动，如下所示。



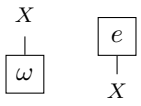
幺半范畴还配有一个特殊的单位对象 I，而 I（的恒等态射）只用空白表示。



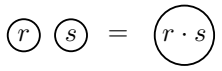
(2)

直观地说，任意对象与 I 作张量积只会使其保持不变。形式上，这通过相干同构 $X \otimes I \simeq X \simeq I \otimes X$ 来表达。

单位对象使我们能够为没有输入和／或输出的态射赋予意义。态射ω: $I \to X$ 称为 X 的一个状态，并用“无输入”来表示。类似地，效应是形如 e: $X \to I$ 的态射，表示为没有输出。

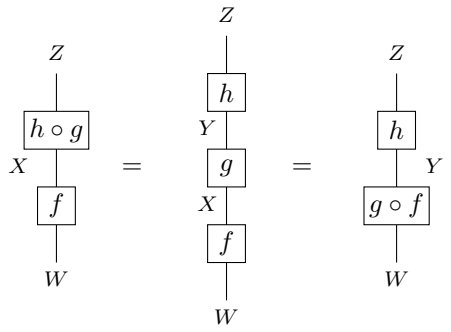


态射 r: $I \to I$ 称为标量，表示为没有输入或输出。标量可以在图示中“自由”移动，也可以通过 $r \cdot s = r \otimes s = r \circ s = s \cdot r$ 相乘，即:

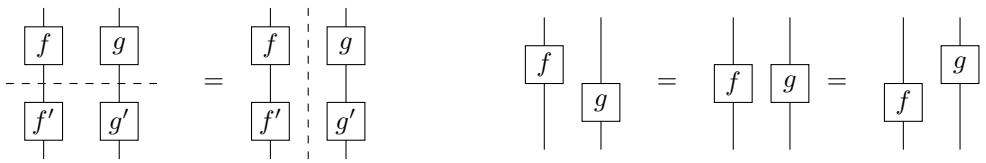


We 将 (2) 中的“空白”标量记为 $1 = \mathrm{id}_I$，它满足：对于所有标量 r，都有 $1 \circ r = r$。

（幺半）范畴中的复合运算 $\circ, \otimes$ 满足许多公理，这里不再赘述；在图示语言中，这些公理显而易见。例如，在任何范畴中，复合的结合律 $(h \circ g) \circ f = h \circ g \circ f = h \circ (g \circ f)$ 在图示中都是自动成立的，如下所示。

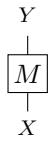


$\otimes$的函子性意味着 $(f \otimes g) \circ (f' \otimes g') = (f \circ f') \otimes (g \circ g')$，这也可从图示（下方左侧）中自动得出。其推论是“交换律” $(f \otimes \mathrm{id}) = f \otimes g = (\mathrm{id} \otimes g)$，根据该定律，我们可以自由地沿着导线滑动方框（下方右侧）。



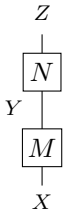
有时，我们常常会像上图那样，省略图示中某些导线（对象）的标签。现在，让我们介绍本文的主要示例范畴。

例 1. 在正矩阵范畴 $\mathbf{Mat}_{\mathbb{R}^+}$ 中，对象是有限集 $X, Y, \ldots$，态射 $M\colon X \to Y$ 是函数 $M\colon X \times Y \to \mathbb{R}^+$，其中 $\mathbb{R}^+ := \{r \in \mathbb{R} \mid r \geq 0\}$。我们将这样的函数视为一个“$X \times Y$矩阵”，其元素为 $M(y \mid x) := M(x, y) \in \mathbb{R}^+$，其中 $x \in X$，$y \in Y$。



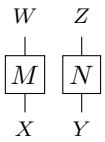
$$:: (x, y) \mapsto M(y \mid x)$$

我们开始合成态射 $N \circ M$，其定义是对内部导线求和 .



$$:: (x, z) \mapsto \sum_{y \in Y} N(z \mid y) M(y \mid x)$$

张量 $\otimes$ 在对象上的定义为 $X \otimes Y = X \times Y$，在态射上的定义则为克罗内克积



$$:: ((x, y), (w, z)) \mapsto M(w \mid x) N(z \mid y)$$

对称性就是显然的同构 $X \times Y \simeq Y \times X$。单位对象是单元素集$I = \{\star\}$。于是，$X$ 的一个状态 ω 等价于 X上的一个正函数：



$$:: x \mapsto \omega(x)$$

其中 $\omega(x) := \omega(x \mid \star)$。后文将讨论的特殊情况，是关于 X 的概率分布。同样，X 上的效应 e 也等价于 X 上的一个正函数，通过 $e(x) := e(\star \mid x)$ 表示。



$$:: x \mapsto e(x)$$

标量 $r\colon I \to I$ 恰好是一个正实数 $r \in \mathbb{R}^+$，而标量的复合相当于在 $\mathbb{R}^+$ 中进行乘法$r \otimes s = r \circ s = r \cdot s$。

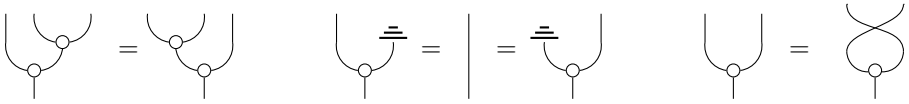
2.2 cd-范畴

诸如 $\mathbf{Mat}_{\mathbb{R}^+}$ 这样的范畴具有更多结构，使我们能够完全以图示方式描述概率理论的诸多方面 [8, 81]。

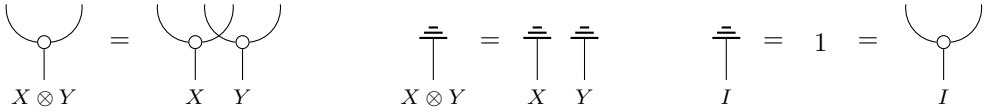
定义 2. *[8]* cd-范畴（复制-丢弃范畴）是一种对称幺半范畴，其中每个对象都配有一对指定的态射



分别称为复制和丢弃，并满足以下条件：

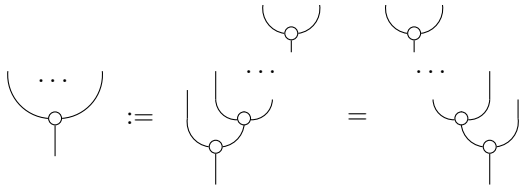


形式上，这些公理表明，复制和丢弃构成一个交换余幺半群。此外，这些态射的选择是“自然的”，因为以下性质对所有对象 X、Y 都成立。



(3)

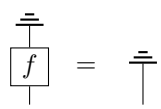
得益于复制公理，我们可以为任意 $n \geq 1$ 明确地定义具有 n 个输出端的复制态射，如下所示。



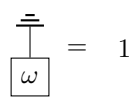
形式上，我们 还将具有 $n = 0$ 个输出端的复制态射定义为丢弃效应 .

在cd-范畴中， h这些过程的 f 一个真ly ‘probabilistic’ or ‘stochastic’ 正是 h些满足i 以下条件。

定义 3. 态射 f 在保持丢弃时称为信道：

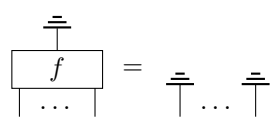


特别地，当状态 ω 满足以下条件时，我们称其为归一化状态。

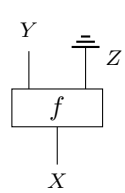


每个态射都是信道的 cd-范畴，或者等价地说，任意对象上的唯一效应都是马尔可夫范畴 *[81]*。给定任意 cd-范畴$\mathbf{C}$，其由信道构成的子范畴$\mathbf{C}_{\text{channel}}$ 总是构成一个马尔可夫范畴。

A 一个由 (3) 得出的有用规则是，任何具有多个输入的信道 f 都满足 :



有了丢弃，就可以“忽略”态射的某些输出。给定任意态射 f 从X 到 Y, Z，其边缘分布 $X \to Y$ 是如下态射。

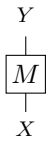


例4. $\mathbf{Mat}_{\mathbb{R}^+}$ 是一个 cd-范畴。在 X 上的复制映射定义为 $(y, z \mid x) = \delta_{x,y,z}$，其取值为 1 当且仅当$x = y = z$，否则为 0。在 X 上的丢弃由满足以下条件的函数给出：$x \mapsto 1$ 对所有 $x \in X$。状态 ω是归一化的，当且仅当它在 X上构成一个归一化概率分布。



$$normalised \iff \sum_{x \in X} \omega(x) = 1$$

更一般地，$M\colon X \to Y$ 是一个信道，当且仅当它是概率通道；等价地说，该矩阵是随机矩阵，意味着它将每个 $x \in X$ 映射为归一化分布。

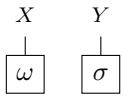


$$a\ channel \iff \sum_{y \in Y} M(y \mid x) = 1 \quad \forall x \in X.$$

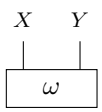
从 X 到 Y 的这类概率通道通常也称为条件概率分布，并记作 P或 $P(Y \mid X)$，其取值记为 $P(y \mid x) := P(Y = y \mid X = x)$，其中 $x \in X$, $y \in Y$。

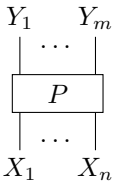
$\mathbf{Mat}_{\mathbb{R}^+}$ 中的信道子范畴是马尔可夫范畴 FStoch，即有限随机矩阵组成的范畴，其中包含有限集之间的概率通道。我们通常会将讨论范围限制在这个子范畴内。

图示现在可以帮助我们表达概率论的一些基本特征。例如，给定 X, Y上的概率分布 ω, σ，它们的张量对应于 $X \times Y$上的独立分布。

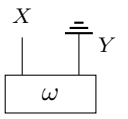


$X \otimes Y$ 上的任意归一化态对应于 X, Y 上的联合分布（如下左图）。类似地，如下右图所示的信道表示一个概率通道 $P(Y_1, \ldots, Y_m \mid X_1, \ldots, X_n)$。

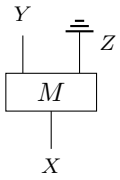




对于 $\mathbf{Mat}_{\mathbb{R}^+}$ 中任意定义在 X 和 Y 上的状态，在通常的概率意义下，对 X 取边缘分布等价于对 Y 求和。一般而言，任意态射的边缘化都通过对被丢弃的对象求和来实现。

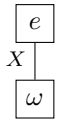


$$:: x \mapsto \sum_{y \in Y} \omega(x, y)$$



$$:: (x, y) \mapsto \sum_{z \in Z} M(y, z \mid x)$$

最后注意，对于任意效应 $e\colon X \to \mathbb{R}^+$ 和归一化态 ω，标量 $e \circ \omega$ 对应于根据概率分布 ω 得到的 e 的期望值。



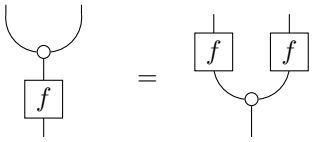
$$= \sum_{x \in X} e(x)\omega(x)$$

在定义因果模型时，我们主要关注信道，因此将在有限概率过程的马尔可夫范畴 FStoch 中开展工作；FStoch 是 $\mathbf{Mat}_{\mathbb{R}^+}$ 中的信道子范畴。不过，对于因果推理（例如第7节和第8节中的计算），在一个更广泛的范畴中开展工作可能会有所帮助，尤其是其中包含非平凡效应的范畴，例如 $\mathbf{Mat}_{\mathbb{R}^+}$。

2.3 确定性过程与帽

在一般的、具有“概率性”特征的过程中，找出那些行为类似于（偏）函数、表现为“确定性”的过程会有所帮助。事实上，这些过程恰好就是保持复制的过程，如下所述。

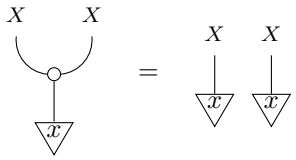
定义5. 若态射 f 满足以下条件，则称其为确定性的。



实际应用中的一个重要特例是确定性状态，我们在此也称之为尖锐态。我们将 X 的尖锐态 x 描绘为：



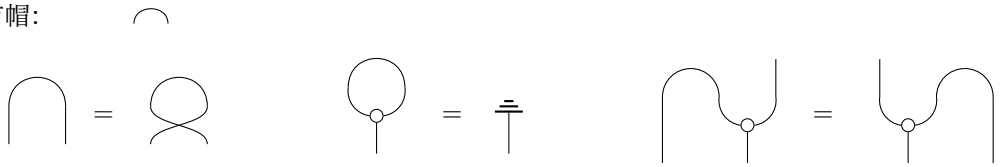
根据定义，这些状态可由复制态射直接复制：



(4)

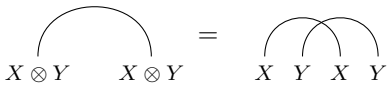
我们将在后续因果推理中发现，确定性状态与效应之间的对应关系十分有用。为此，我们要求 cd-范畴具有一个额外性质。

定义6. 若 cd-范畴 C 中的每个对象 X 都配有一个指定的、作用于 $X \otimes X$ 的效应（描绘为 ），并满足以下条件，则称 C 具有帽：

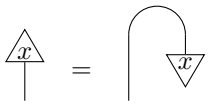


(5)

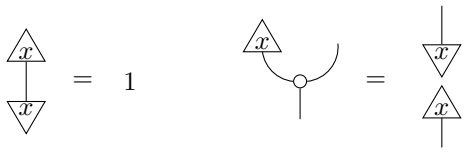
此外，帽的选取是“自然的”，因为对于所有对象 X, Y，以下性质都成立。



由此可知，每个帽都是确定性的。[4] 现在，在任何带帽的 cd-范畴中，对于任意尖锐态 x，都有一个相应的确定性效应，记作 $x^\dagger$，并描绘为“将 x 上下翻转”：



若一个效应具有 $x^\dagger$ 的形式，我们就称它为尖锐效应。可以验证，对于任意尖锐态 x，效应 $x^\dagger$ 是满足以下条件的唯一效应。

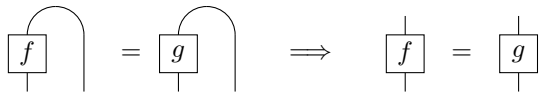


(6)

[4]可以验证，式（5）中的右侧态射满足交换律和结合律，从而在 C 中构成一种交换的 $semigroup$ 乘法。

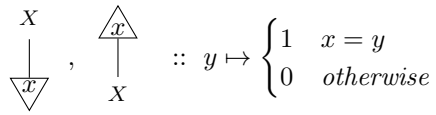


帽满足　当它们是可消去的时，在图示推理中尤其有用，也就是说　　t:



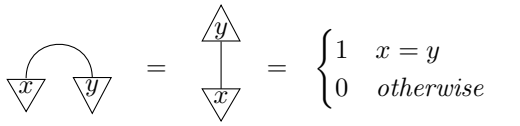
对所有态射 f, g,

例 7。$\mathsf{Mat}_{\mathbb{R}^+}$ 具有可消去帽。每个点 $x \in X$ 都对应于 X上的一个归一化尖锐态，我们仍记作 x，它由 X 上的点概率分布 δ_x 给出。尖锐效应 $x^\dagger$ 也由 X 上的同一个函数 δ_x 给出。



$$:: \; y \mapsto \begin{cases} 1 & x = y \\ 0 & otherwise \end{cases}$$

帽由 $(x, y) = \delta_{x,\ y}$ 给出。因此，对于所有 $x, y \in X$，以下结论成立。



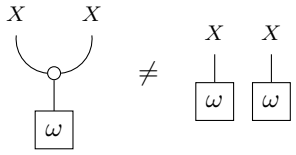
$$= \begin{cases} 1 & x = y \\ 0 & otherwise \end{cases}$$

X 上的每个尖锐态都具有上述形式，其中某个 $x \in X$；否则，它由零态 0 给出，该零态定义为：对所有 $x \in X$, $0(x) = 0$。那么，$0^\dagger$ 也由 X 上的常零函数给出。特别地，唯一的尖锐标量是 0 和 1。我们注意到一个有用的观察：对于任意态射 $M\colon X \to Y$，其标量值$M(y \mid x) \in \mathbb{R}^+$ 可以通过如下方式与 x 和 $y^\dagger$ 复合，以图示地表示。

$$M(y \mid x) \quad =$$

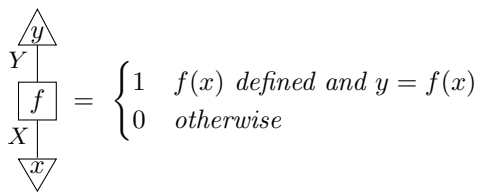


由于 $(y, z \mid x) = \delta_{x,y,z}$ 确实会按照（4）所示复制由 $x \in X$ 给出的每个状态。相比之下，一般态 ω 不可复制。



上面的左侧为 $(x, y) \mapsto \omega(x)^{\delta_{x,y}}$，而右侧为 $(x, y) \mapsto \omega(x)\omega(y)$；除非 ω 为零，或对于某个 $x \in X$有 $\omega = x$，否则两者并不相同。左侧定义了一个分布，其中X 的副本完全相关；右侧则是一个乘积分布，其中两个 X 彼此独立。

$\mathsf{Mat}_{\mathbb{R}^+}$ 中的确定性态射对应于（部分）函数。任意部分函数 $f\colon X \to Y$定义了一个态射 $X \to Y$，我们仍将其记为 f；它在尖锐态 $x \in X$上的取值为 $f \circ x = f(x)$，当且仅当 $f(x)$有定义时；否则为 $f \circ x = 0$。等价地，



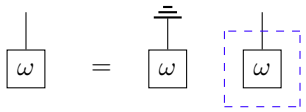
$$= \begin{cases} 1 & f(x) \text{ defined and } y = f(x) \\ 0 & otherwise \end{cases}$$

cd-范畴的框架远比 $\mathsf{Mat}_{\mathbb{R}^+}$ 的有限离散情形更加一般。在附录A中，我们详细介绍了一种用于处理完整测度论概率的 cd-范畴，不过阅读本文其余部分并不需要了解这一内容。

本节余下部分将介绍 cd-范畴的更多性质，这些性质对因果推理很有用。虽然它们并非定义因果模型本身所必需，但在第7节和第8节中将非常有用。

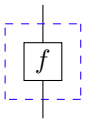
2.4 归一化

在因果推理中，通常需要对状态和过程进行归一化。直观地说，对于一般态 ω，其归一化应是另一个状态 $\mathrm{norm}(\omega)$，用围绕 ω 的虚线方框表示，并满足以下条件：



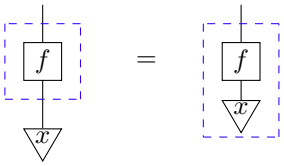
(7)

在 $\mathrm{Mat}_{\mathbb{R}^+}$ 中，这对应于将一个状态写成一个非负常数与一个“正确归一化”的概率分布的积。更一般地，对于任意过程 f，其归一化范数 $\mathrm{norm}(f)$ 是如下图示的过程：



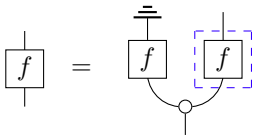
(8)

直观地说，对于每个尖锐态输入 x，它都是通过对 $f \circ x$ 进行归一化得到的。



(9)

还有一种不涉及状态、表达后者的方法，如下所示。

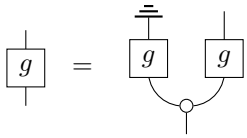


(10)

将尖锐态 x 代入 (10)，可知 (7) 对 $\omega = f \circ x$ 成立，其中 $\mathrm{norm}(\omega) = \mathrm{norm}(f) \circ x$。当 $f = \omega$ 已经是一个状态时，这就简化为 (7)。

通常可以将归一化过程视为一个信道，因此对于状态 ω，其归一化 $\mathrm{norm}(\omega)$ 按照我们先前的定义是归一化的。但这并非总是可行。例如，在 $\mathrm{Mat}_{\mathbb{R}^+}$ 中，如果 ω 是零态，那么 $\mathrm{norm}(\omega) = 0$。更一般地，如果对于某个输入 x，$f(x) = 0$，那么 $\mathrm{norm}(f)(x) = 0$ 也如此。

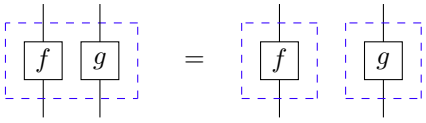
相反，归一化 $g = \mathrm{norm}(f)$ 通常只会是一个部分信道，这意味着以下结论成立。



现在，我们可以正式定义归一化，并陈述它的一些基本性质 s.

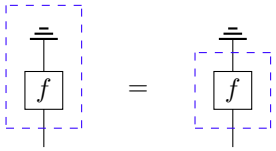
定义 8。设 C 为一个 cd-范畴。如果对于每个态射 f，都存在一个指定的部分信道 $\mathrm{norm}(f)$，如式 (8) 所示，以虚线方框表示，并且满足 (10) 以及以下条件，则称 C 具有归一化：

1. 若 f 是一个部分信道，则有 $\mathrm{norm}(f) = f$。
2. 对所有态射 f, g：



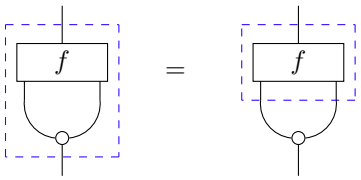
(11)

3. 对所有态射 f：



(12)

4. 对所有态射 f:

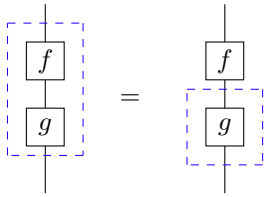


(13)

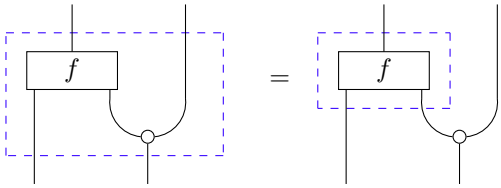
在附录 B 中，我们实际上证明了，只要 cd-范畴 C 上存在归一化结构，该结构便是唯一的，因为上述公理完全确定了它。我们还在附录 B 中确立了进一步的条件，这些条件保证 cd-范畴具有归一化，并给出了归一化的以下进一步性质。

引理 9. 设 C 是具有归一化的 cd-范畴。

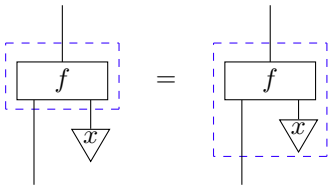
4. 当 f 是信道时，对于所有态射 g，我们有:



5. 对于所有态射 f，我们有:



6. 若 C 有帽，则 (9) 成立；更一般地，对于每个态射 f 和尖锐态 x，我们有:



(14)

例 10. $\mathrm{Mat}_{\mathbb{R}^+}$ 具有归一化。对每个对象 X，由 0 给出的零态 $\mathbf{0}(x) = 0$ （对所有 $x \in X$）满足 $\mathrm{norm}(\mathbf{0}) = 0$。对于非零态 ω，我们有

$$\mathrm{norm}(\omega)(x) = \frac{\omega(x)}{\sum_{x' \in X} \omega(x')}$$

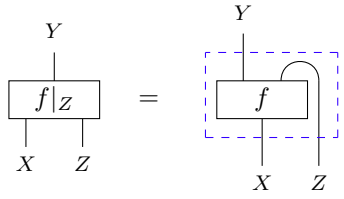
对于一般态射 $M\colon\ X \to Y$，归一化 $\mathrm{norm}(M)$ 将 $x \in X$ 发送到 $\mathrm{norm}(M(x))$。也就是说:

$$\mathrm{norm}(M)(y \mid x) = \begin{cases} \frac{M(y|x)}{\sum_{y' \in Y} M(y'|x)} & \text{if } \sum_{y \in Y} M(y \mid x) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

2.5 条件化与条件独立性

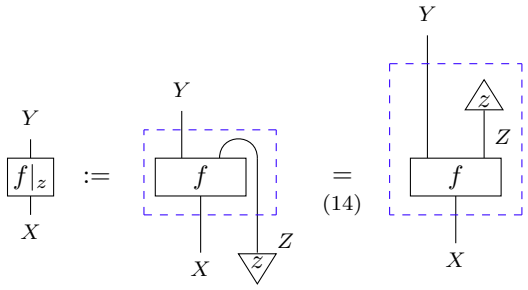
就我们的目的而言，归一化的主要用途是定义概率条件化。在这里，我们给出一种条件化的图示化论述；在具备归一化和良定义的帽的情况下，它将 Cho 和 Jacobs [8] 以及 Fritz[10] 此前关于信道的处理扩展到了一般过程。

定义 11. 如果一个 cd-范畴 C 具有归一化和可消去帽，我们就称其具有效应条件化。于是，对于任意态射 $f\colon X \to Y \otimes Z$，我们将条件分布 $f|_Z\colon X \otimes Z \to Y$定义为部分信道



(15)

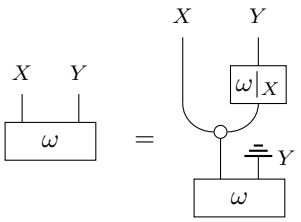
*For an*y 的确定性状态 z属于 Z，随后我们通过与 z复合来定义 $f|_z\colon X \to Y$，使得



(16)

下面这个有用性质推广了概率论中的链式法则或乘法法则 $P(x,y) = P(y|x)P(x)$；对于 ω 已归一化的情形，Cho 和 Jacobs 将其称为联合态 [8]的分解。

Le理 12（分解）。如果 C 具有效应条件化，那么 $X \otimes Y$ 的任意状态 ω 都满足



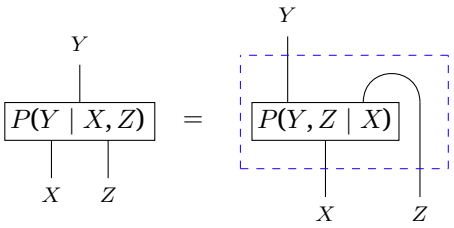
(17)

我们在附录 B 中证明了这一结果的一个扩展，说明我们的定义 $f|_Z$ 在何种意义上满足参考文献 [8, 10]中的“条件分布”的范畴论定义。请注意，只有存在可消去帽时，这一点才成立。后面这些方法对条件化（针对信道）给出了更一般的说明，但本文不采用这些方法。

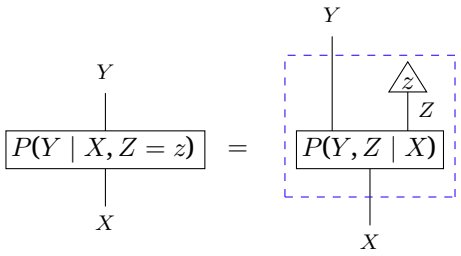
例 13。正如我们所见，$\mathbf{Mat}_{\mathbb{R}^+}$ 具有效应条件化。对于态射 $M\colon X \to Y \otimes Z$，部分信道 $M|_Z$ 由下式给出

$$M|_Z(y \mid x,z) = \frac{M(y,z \mid x)}{\sum_{z' \in Z} M(y,z' \mid x)}$$

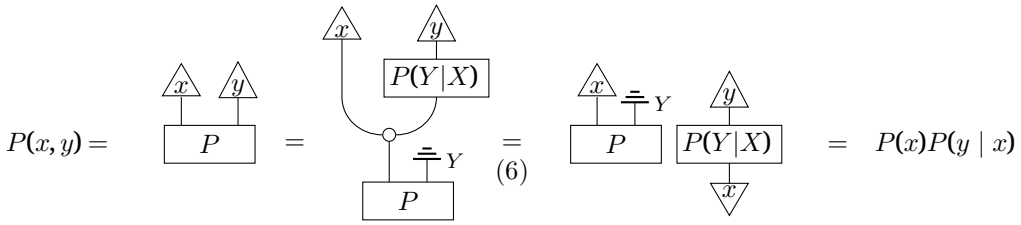
只要分母中的和不为零，且 $M|_Z(y \mid x,z) = 0$对所有 y均成立。按照标准概率论记法，信道 $M\colon X \to Y \otimes Z$由密度 $P(Y,Z \mid X)$描述，而 $M|_Z$对应条件密度 $P(Y \mid X,Z)$。



对于任意 $z \in Z$，部分信道 $M|_z$ 对应条件密度 $P(Y \mid X, Z = z)$。



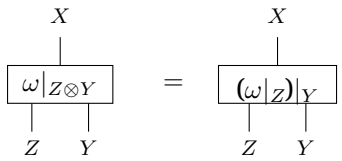
要看出 (17) 与概率论中乘法法则的关系，请注意，对于任意联合分布$P(X, Y)$，我们有如下结果。



注释14. 在归一化的定义中引入部分信道，意义在于它使我们能够为没有全支撑的分布定义条件分布；这类分布对于因果推理的诸多方面都很重要。与上文一样，在 $\mathsf{Mat}_{\mathbb{R}^+}$ 中，条件分布 $P(X|Z = z)$ 会在 $P(Z = z) = 0$ 时输出零态，将其解释为 X 上的“未定义”分布；否则则给出预期的条件分布。

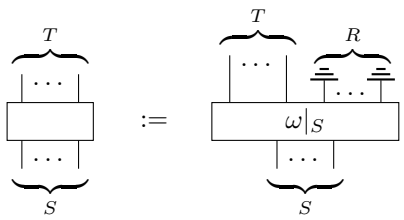
现在我们将证明条件分布的行为符合我们的直觉，所有证明见附录C。

命题1 5. 设 C 为一个带有效应条件化的 cd-范畴，且 ω 是 $X \otimes Y \otimes Z$ 的一个状态。该 n



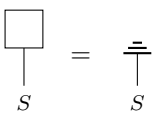
(18)

这与（12）结合起来意味着，相对于给定状态 ω，可以省略条件分布方框中的标签——其身份由输入和输出类型唯一确定。只要上下文能清楚表明所指的参考状态 ω，采用以下记法便很方便。给定 $X_1 \otimes \cdots \otimes X_n$ 的状态 ω，对于任意不相交子集 S、$T \subseteq \{X_1, ..., X_n\}$ 和 $R := \{X_1, ..., X_n\} \setminus (S \cup T)$，我们记作：



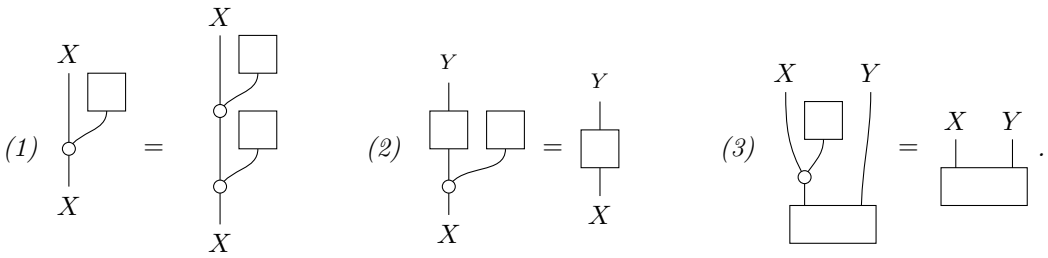
(19)

特别地，当满足以下条件时，我们称 ω 在 $S \subseteq \{X_1, \ldots, X_n\}$ 上具有全支撑：



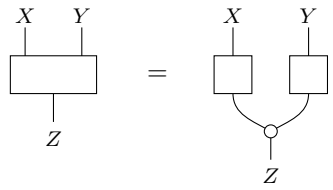
即使 ω 在 S上缺乏全支撑，这个方框仍会在 S 上诱导一个态射，该态射在复合下保持所有条件分布不变，如下所示。

引理 16. 设 C 是一个具有效应条件化的 cd-范畴，且 ω 是 $X \otimes Y$的一个状态。则以下结论成立。



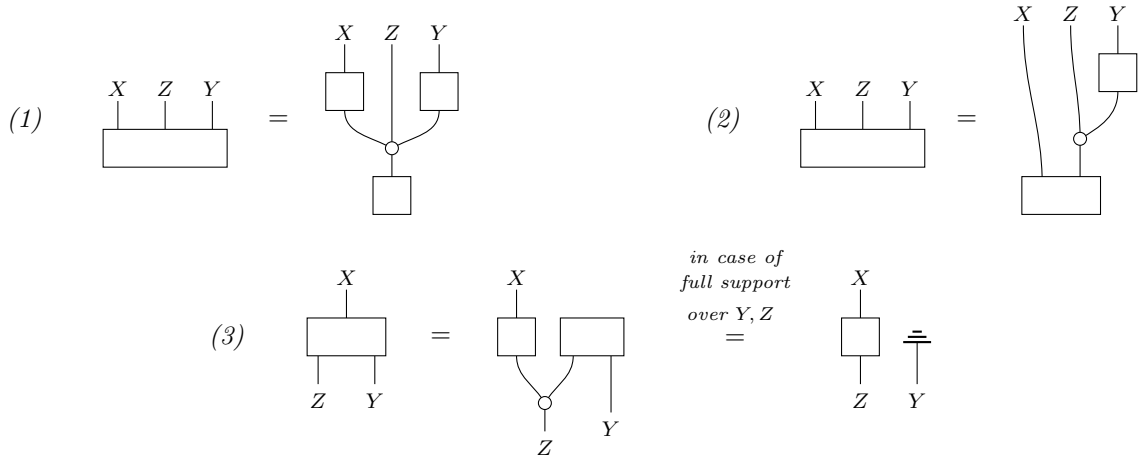
注意， X 当然可以具有如上所述的、基于对象子集 S 的积结构。现在我们可以定义条件独立性，使参考文献 [8, 10] 中的同一思想能够以此处所述的条件分布形式呈现。

定义 17。（条件独立性）：设 C 是带有效应条件化的 cd-范畴。给定 $X \otimes Y \otimes Z$的一个状态ω，若 X 相对于 Z 独立于 Y，记作 $(X \perp\!\!\!\perp Y|Z)\omega$，当且仅当



观察以下等价关系颇有启发。

引理 18。设 C 是带有效应条件化的 cd-范畴。给定 $X \otimes Y \otimes Z$的一个状态 ω，条件独立性 $(X \perp\!\!\!\perp Y|Z)\omega$ 等价于以下条件：



与 [8], 中的全支撑情形类似，我们能够为用于推理条件独立性的经典半拟阵公理建立弦图形式；这些公理在附录 C.4 中回顾并证明。

定理 19. 设 C 是一个具有效应条件化的 cd-范畴，且 ω 是某个状态。定义 17 中的条件独立关系 $(\perp\!\!\!\perp \_\_|\_)\omega$，将其理解为任意三个不相交子集（ ω 的陪域中的对象的子集）上的三元关系，则满足半拟阵公理。

例 20。在 $\mathsf{Mat}_{\mathbb{R}^+}$ 中，根据定义 17，条件 $(X \perp\!\!\!\perp Y|Z)_P$ 对于某个 $P(X,Y,Z)$ 确实表示为

$$P(X,Y|Z) \;=\; P(X|Z)\; P(Y|Z)\;,$$

理解为：当 $P(Z=z) \neq 0$ 时为 $P(X,Y|Z=z) = P(X|Z=z)P(Y|Z=z)$，而对于所有其他 z，在 $\mathbf{Mat}_{\mathbb{R}^+}$ 中将其设为零态。引理 18 的条件 (1) 和 (2) 分别为：

$$P(X,Y,Z) \;=\; P(Y|Z)\; P(X|Z)\; P(Z) \quad \textit{and} \quad P(X,Y,Z) \;=\; P(Y|Z)\; P(X,Z)\;,$$

如上所述，两种情况都涉及对全支撑的依赖。在全支撑情形下，条件 (3) 表示为

$$P(Y|X,Z) \;=\; P(Y|Z)\;.$$

3 因果模型

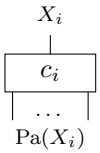
3.1 定义

现在让我们转向本文的主要研究对象，并给出 cd-范畴中因果模型的一般定义。回顾一下，有限有向图 $G = (V, E)$ 由有限的顶点集合 $V = \{X_1, \ldots, X_n\}$ 和有向边子集 $E \subseteq V \times V$ 组成。边 $(X_i, X_j) \in E$ 通常记作 $X_i \to X_j$。此外，有限有向无环图 (DAG) 不含有向环，也就是说，不存在由形如 $X_{i_1} \to X_{i_2} \to \cdots \to X_{i_n} \to X_{i_1}$ 的边构成、长度为 $n \geq 1$ 的序列。对于顶点 $X \in V$，分别用 $\mathrm{Pa}(X) \subseteq V$ 和 $\mathrm{Ch}(X) \subseteq V$ 表示 X 的父节点集合和子节点集合。

因果模型本质上由一个有向无环图指定，其边描述变量之间的直接因果关系；此外，还包括描述有向无环图中每个变量由其父节点决定的机制的信道。

定义 21. 设 $\mathbb{C}$ 为一个 cd-范畴。$\mathbb{C}$ 中的因果模型 $\mathbb{M}$ 由以下内容给出：

1. 一个顶点为 $V = \{X_1, \ldots, X_n\}$ 的有限有向无环图 G；
2. 指定的输出顶点子集 $O \subseteq V$；
3. 对于每个顶点 $X_i \in V$，一个 $\mathbb{C}$ 中与之关联的对象（也记作 X_i），以及 $\mathbb{C}$ 中一个形如……的信道 m:



(20)

其中，$\mathrm{Pa}(X_i)$表示 G中 X_i 的父节点。我们称信道 c_i 为 X_i的机制。

有序对 $G_{\mathbb{M}} = (G, O)$称为 $\mathbb{M}$的因果结构。$\mathbb{C}$ 的指标对象 $\mathrm{var}(\mathbb{M}) = (X_i)_{i=1}^n$ 称为模型的变量。与输出顶点对应的变量称为输出变量，记作 $\mathrm{out}(\mathbb{M}) \subseteq \mathrm{var}(\mathbb{M})$。

请注意，这一定义允许只将变量的一个子集视为模型的“输出”。这一说明不一定意味着这些是唯一“观测到的”变量；它是一项可用于不同目的的数据。

除了直接指定有向无环图（DAG）之外，我们也可以只通过机制的集合来指定因果模型。更准确地说，我们可以等价地将 $\mathbb{C}$ 中的因果模型 $\mathbb{M}$ 定义为由指标化集合 $\mathrm{var}(\mathbb{M}) = (X_i)_{i=1}^n$、一个子集 $\mathrm{out}(\mathbb{M})$以及一组信道 $(c_i)_{i=1}^n$ 构成，如式 (20) 所示；其中，有向图 G 的顶点为 X_i，并且每当 X_i 是 c_j 的输入时，就有一条边 $X_i \to X_j$，且该图无环。

当我们采用该范畴来描述有限集之间的概率映射时，便立即得到更标准的因果模型概念。

例 22。一个因　　因果贝叶斯网络（CBN）是 $\mathbb{C} = \mathbf{FStoch}$ 中的一种因果模型。因此它由以下部分组成：

- *a finite DAG $G = (V, E)$ with a subset $O \subseteq V$ of outputs;*

- 对每个顶点 X_i，都有一个与之关联的变量，其取值构成有限集合，也记作 X_i；还有一个机制c_i，由具有密度的概率通道给出

$$P(X_i \mid \mathrm{Pa}(X_i)) \tag{21}$$

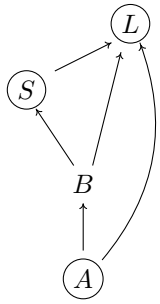
因果贝叶斯网络在所有变量上定义了一个联合分布，其密度为

$$P(X_1, \ldots, X_n) := \prod_{i=1}^{n} P(X_i \mid \mathrm{Pa}(X_i)) \tag{22}$$

正如符号所示，这些机制通常可以视为关于 $P(X_1, \ldots, X_n)$ 的条件密度，不过我们要强调，它们是以信道的形式给出的——参见注 23。一般而言，若存在概率信道使得式 (22) 成立，则称任意联合密度 P 满足关于 G 的因果马尔可夫条件。人们通常尤其关注因果贝叶斯网络的输出分布，即输出变量 O 上的边缘分布。

注 23。更常见的做法或许是将因果贝叶斯网络定义为由有向无环图 G 及其对应变量，以及满足马尔可夫条件 (22) 的联合分布 $P(X_1, \ldots, X_n)$ 给出，并将机制定义为相应的条件密度 (21)。然而，如果联合分布不具有全支撑，这些条件信道便无法被唯一确定；在考虑干预（第 4 节）等情形时，具体选择可能会产生实际影响。相比之下，我们认为特定的信道 c_i 才是因果模型的关键，因为它们具有作为因果机制的解释。因此，我们将因果贝叶斯网络定义为由其机制而非联合分布所确定。

例 24. 作为因果贝叶斯网络的一个基础且更具体的例子（下文也将贯穿使用），来看下面这个例子，取自参考文献 *[37]*。与第 1.1 节中的例子类似，它讨论的是吸烟和肺癌，不过这次还包含其他变量，这些变量将在后文研究政策制定时发挥作用。设 S 为某人是否选择吸烟，L 为其是否罹患肺癌，A 为其年龄，B 为一组相关的背景条件，如社会经济地位、教育等。一个合理的因果模型由下方的有向无环图构成，其中对应输出变量的顶点用圆圈标出；此外，右侧还列出了各个概率通道的具体定义，它们给出了相应的机制。



$P(L|SBA)$

$P(S|B)$

$P(B|A)$

$P(A)$

正如注释 23 所强调的，因果贝叶斯网络并非由联合分布所确定的条件分布定义，而是由因果机制定义。不过，这些数据当然定义了输出变量上的边缘分布：

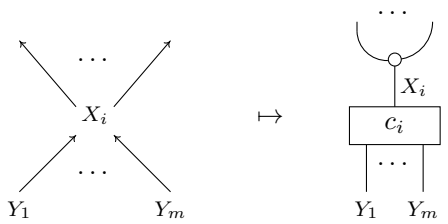
$$P(SLA) = \sum_{B} P(L|SBA)\ P(S|B)\ P(B|A)\ P(A)$$

3.2 作为图示的因果模型

现在让我们看看如何用图示描述因果贝叶斯网络以及更一般的因果模型。事实上，有向无环图与某一类特定的弦图之间存在精确对应关系，这种对应关系是因果模型图示方法的核心。这一构造最初由 Fong [7] 提出，随后 Jacobs 等人 [9]对其作了更明确的阐述。

定义 25. 设 G 是顶点为 $V = \{X_1, \ldots, X_n\}$的有向无环图， $O \subseteq V$是输出顶点的一个子集。我们定义一个弦图 $D_{G,\ O}$，它没有输入，并且对每个 $X_i \in O$都有一个输出，定义如下。

1. 对每个顶点 X_i，我们从 X_i 的父节点 Y_1、.....、Y_m 到 X_i 画一个方框，记作 c_i，如下所示。接着，如果 X_i 是输出变量，我们就将方框 $k+1$ 的输出复制 $k+1$ 次；否则复制 k 次，其中 k 是 X_i 的子节点数。

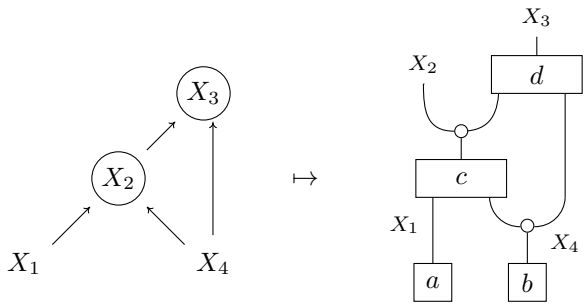


特别地，如果变量 X_i 没有父节点，那么方框 c_i 没有输入；如果 X_i 没有子节点且不是输出变量，那么 c_i 前面会接上丢弃效应 。

2. 然后，我们通过将一个 c_i 的输出（后接复制映射）接入 c_j 的 X_i 输入，将这些部分连接起来；每当 G 中存在一条边 $X_i \to X_j$ 时，就进行这样的连接。

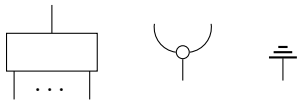
由于 G 是无环的，这就得到了一个类型明确的字符串图。特别地，按照构造，只有当 $X_i \in O$ 时，每个顶点 X_i 才会作为输出出现，而且此时恰好出现一次。实际上，我们允许为图中的方框 c_i 任意命名。

例26. 对于下方左侧的有向无环图 G （输出以 $O = \{X_2, X_3\}$ 圈出），其图 $D_{G,O}$ 如右侧所示。



根据构造，每个弦图$D_{G,O}$ 只包含单输出盒子和复制映射。更准确地说，我们可以如下识别以这种方式从有向无环图得到的弦图类型。

定义27. 网络图是由单输出盒子、复制映射和丢弃效应构成的弦图 D ：



并在线上附加标记，使得任何未通过一系列复制映射相连的线都具有不同的标记，且每个标记至多出现一次作为输出，并且至多出现一次作为任意给定方框的输入。状态网络图是不含输入的网络图。

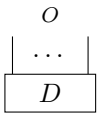
迄今定义的因果模型只需要状态网络图，但带输入的一般网络图将在第5节以及第7节和第8节中使用。

当两个网络图根据 cd-范畴的一般公理相等（允许对线和方框进行重新标记）时，我们认为它们等价。

备注 28. 网络图是 Fritz 和 Liang *[35],* 所研究的“gs-幺半范畴字符串图”的一个子集，后者允许方框具有多个输出。这里不需要考虑后者，因为因果模型中的每个变量都是其机制的唯一输出。

Prop命题 29. 在等价以及对图重新标记的意义下，以下对象等价 t:

- 一个有向无环图 $G=(V,E)$，其输出为 $O\subseteq V$;
- 一个状态网络图



带有导线 V 和输出 O;

通过 $(G,O)\mapsto D_{G,\ O}$ 以及 $D\mapsto(G_D,O_D)$。

证明。给定 (G,O)，我们如定义25中所述定义 $D_{G,\ O}$。反过来，给定一个状态网络图 D，我们定义有向无环图 G_D：对于 D 中每个非复制方框的输出，都设一个顶点 X；若 $X\in O_D$ 是该图的一个输出，则令其成立。每当 X 是唯一一个输出为 Y 的方框的输入时，我们就加入一条有向边 $X\to Y$。忽略重新标记后，很容易看出这些映射互为逆映射。□

目前为止，这种对应关系将有向无环图与状态网络图联系起来；它们本质上纯属句法描述，只勾勒出因果结构。要给出一个实际的因果模型，我们必须在某个范畴中对它们进行解释，将导线和方框分别对应到该范畴中的特定对象和信道。

定义 30. cd-范畴 C 中网络图 D 的解释 $[\![-]\!]$，包括在 C 中为每条导线 X_i 指定一个对象 $[\![X_i]\!]$，并为每个方框 f 指定一个信道 $[\![f]\!]$: $[\![X_1]\!]\otimes\cdots\otimes[\![X_k]\!]\to[\![X]\!]$ 对于 D 中输入为 X_1、 ..... 、 X_k 且输出为 X 的每个方框。

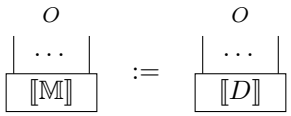
因此，对 D 的解释 $[\![-]\!]$ 涉及为其中D的所有导线指定对象，并为所有方框指定信道。为简便起见，有时我们也把网络图 D 连同其在 C 中的解释 $[\![-]\!]$，称为 C 中的网络图 D。

定理 31. 设 C 是一个 cd-范畴。在 C 中指定一个因果模型 $\mathbb{M}$，等价于指定一个状态网络图 D 及其在 C 中的解释 $[\![-]\!]$。

证明. 根据命题 29，因果结构（G,O）等价于状态网络图 $D=D_{G,O}$。随后，对网络图 D 进行解释，恰好等同于为每个变量指定一个机制 (20)，也就是指定一个具有该因果结构的因果模型。□

简言之，C 中的因果模型 $\mathbb{M}$ 与状态网络图相同。这样，只需绘制模型的网络图，即可指定一个完整的因果模型；网络图编码了因果结构 (G,O) 并标明各个机制，然后再在 C 中解释这些机制即可。在实践中，尤其是在第7节和第8节中，给定模型 $\mathbb{M}$ 时，我们通常省略书写 $[\![-]\!]$，因此对于网络图中的标签和方框及其在 C 中对应的解释，都只写作 X, f，而不将后者记为 $[\![X]\!]$,$[\![f]\!]$。

模型定义的输出态。给定任意网络图 D 及其在 C 中的解释 $[\![-]\!]$，我们可以按照 D 合成图中的解释后的信道，从而得到 C 中一个从图的输入到输出的总体通道，记为 $[\![D]\!]$。特别地，对于任意因果模型 $\mathbb{M}$，其网络图为 D 且输出为 O，我们可以得到 C 中定义在这些输出 O 上的归一化态:



我们称之为模型的输出态。

反过来，在实践中，我们通常从一组对象 O 上的归一化态 ω 出发，并询问 ω 是否与某个有向无环图 G 所给出的因果结构兼容，或等价地，是否与某个网络图兼容。若存在对 D 的一种解释 $[\![-]\!]$，其输出

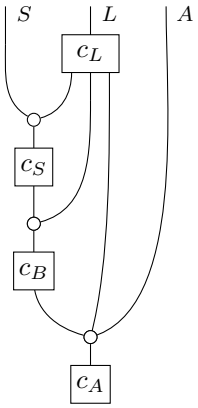
对象 O 与 ω 的相同（即为图中其余非输出线选择对象，并为图中的所有方框选择信道），使得 $[\![D]\!] = \omega$，则称 G 按照网络图 V 进行因子分解。同样，若 ω 按照具有指定输出 $O \subseteq V$ 的有向无环图 G 在 V 上进行因子分解，且它按照 $D_{G,\ O}$ 进行因子分解，则称其按照该有向无环图进行因子分解。

例32。在 FStoch 中解释网络图，需要为每个标签 X_i 指定有限的取值集合，并为图中每个方框指定一个概率通道 $c_i\colon Y_1 \times \cdots \times Y_n \to X_i$，其输入和输出与该方框相同。因此，我们可以通过网络图及其各方框作为有限概率通道的解释来指定一个因果贝叶斯网络，简言之，就是 FStoch 中的网络图。

给定这样的因果贝叶斯网络 $\mathbb{M}$，C 中的输出态 $[\![\mathbb{M}]\!]$ over O 描述了输出变量 O 上的结果输出分布 $P(O)$。

一般来说，只要 ω 满足网络图 $D = D_{G,O}$ 的因果马尔可夫条件（见例22），分布 $\omega = P(O)$ 就会按该网络图因子分解。

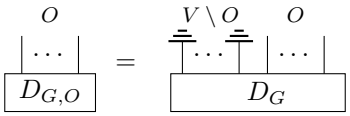
例33。回顾例24中的因果贝叶斯网络。它的网络图表示 D 由下图给出，并将 c_L、c_S、c_B 和 c_A 分别解释为信道 $P(L|SBA)$、$P(S|B)$、$P(B|A)$ 和 $P(A)$。



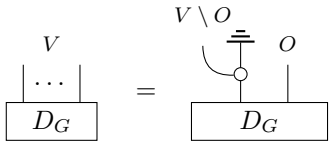
(23)

将上图视为 S、L、A 上的单个状态，恰好得到该模型的输出态，即输出分布 $P(S, L, A)$。

值得指出的是，给定一个与有向无环图 G 对应的状态网络图$D = D_{G,\ O}$，其顶点为 V，输出为 O，我们总可以将其扩展为图 $D_G := D_{G,\ V}$，它基于同一个有向无环图，但此时每条导线都作为输出。事实上，$D_{G,\ O}$ 可以作为后者的边缘分布得到：

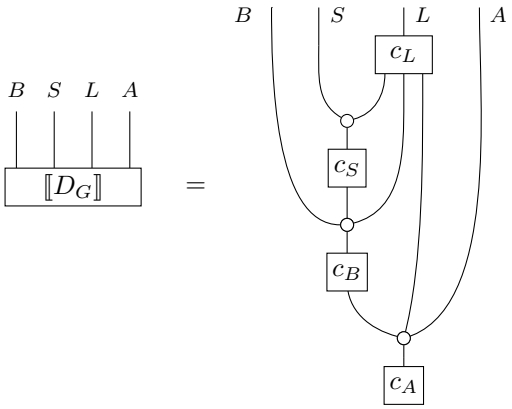


反过来，我们也可以通过将每条非输出线额外复制一次来得到 D_G：



例34。对于因果贝叶斯网络，归一化态 $[\![D_G]\!]$ over V in C 描述的是模型全部变量集合 V 上的概率分布 $P(V)$，而不仅仅是输出 O。对于图如下的因果贝叶斯网络

D 在变量为 $V = \{A, B, S, L\}$ 的例（23）中，这由下式给出：



备注35。如 *[9],* 所示，网络图的解释 D 可以等价地定义为一个 cd-函子

$$\mathbf{Free}(D) \to \mathbf{C}$$

其中 Free(D) 是由图中的标签和方框生成的自由 cd-范畴。因此，DAG G 上的范畴 C 中的因果模型由（其输出 O 以及）一个 cd-函子 Free(D_G) $\to$ C 决定。这是 *[7],* 中关于因果模型的原始观点，Fong 将范畴 Free(D_G) 称为 G的因果理论。这种观点与我们的方法相容，但在这里，我们不太强调将模型视为函子的观点。

3.3 函数因果模型

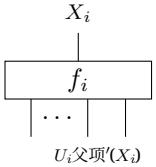
在 Pearl 因果模型框架中，除了因果贝叶斯网络（CBN）之外，还有一个因果模型的核心概念，即函数因果模型（FCM），也称为结构因果模型（SCM）。这种模型被认为编码了更为“精细”的因果知识，是更详细的因果解释以及回答不同类型因果查询的基础，尤其是反事实查询——它是与因果层级 [4, 14]中第三层、也是最高层相关联的模型。

在结构因果模型中，每个机制都分解为一个确定性函数组件（表示底层因果过程）和一个额外的、没有父节点的噪声变量，后者编码了我们对世界底层状态的不确定性。从概念上看，结构因果模型可以视为因果贝叶斯网络的基础：前者基于定义为函数依赖的因果关系来形式化因果推理，随后通过对“局部噪声变量”进行边缘化，得到后者。不过，从形式上看，结构因果模型可以视为定义在更大变量集上的因果贝叶斯网络的一个特例，以下论述将利用这一事实。

在定义5中，我们看到函数在 cd-范畴中可以推广为确定性信道。由此自然得到以下定义。在我们的范畴论语境中，我们倾向于使用“函数因果模型”这一术语，而将“结构因果模型”留给它们在 FStoch 中的通常实例。

定义36。设 C 为一个 cd-范畴。C 中的函数因果模型（FCM）是一种因果模型 $\mathbb{M}$，其变量被划分为 $V = \{X_i\}_{i=1}^n$（称为内生变量）和 $U = \{U_i\}_{i=1}^n$（称为外生变量），输出变量为 $O \subseteq V$，并且满足：

1. 每个 U_i 都没有父节点；2. 每个 X_i 都有如下形式的机制



(24)

其中 $\mathrm{Pa}'(X_i) \subseteq V$ 且 f_i 是一个确定性信道。

也就是说，FCM 是一种因果模型，其中每个内生变量 X_i ——即模型所建模或解释的变量——都有一个确定性机制 f_i，该机制包含一个非输出父节点 U_i；而每个这样的 U_i 变量——被视为“局部扰动”——都没有父节点，且只有一个子节点。由于每个 U_i 都没有父节点，根据因果模型的定义，其机制将只是一个状态



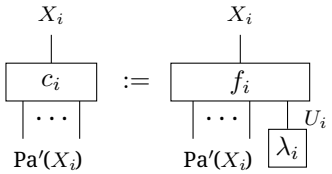
文献中所研究的通常的函数因果模型类型如下。

示例 37. 结构因果模型（SCM）是 FStoch 中的一个函数因果模型。因此，它由有限集X_1、.....、X_n、U_1、.....、U_n给出，分别描述内生变量和外生变量的取值，并且在每个 U_i 上都有一个分布 λ_i，且对于每个 X_i 都有一个函数

$$f_i \colon \mathrm{Pa}'(X_i) \times U_i \to X_i$$

对于子集 $\mathrm{Pa}'(X_i) \subseteq \{X_1, \ldots, X_n\}$。

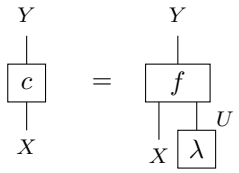
通常认为，具有变量 $V \cup U$ 的函数因果模型表示支撑（非函数型）因果模型的确定性机制 f_i，其中所有随机性都来自我们对 U_i的无知，这种无知由“噪声”分布 λ_i刻画。给定任意函数因果模型 $\mathbb{M}$，我们可以通过为每个 X_i 赋予机制，将因果模型 $\mathbb{M}|_V$ 定义在仅包含 V的变量上



(25)

反过来，FStoch 中的每个因果模型，也就是每个因果贝叶斯网络，实际上都具有某个结构因果模型 $\mathbb{M}$ 的这种形式，因为借助以下众所周知的事实，我们总能将任意概率机制扩展为函数部分，并附加一个“噪声”分布。

命题 38. 对于 Stoch 中的每个信道 c，都存在一个对象 U，它具有一个状态 λ 和一个确定性信道 f，使得

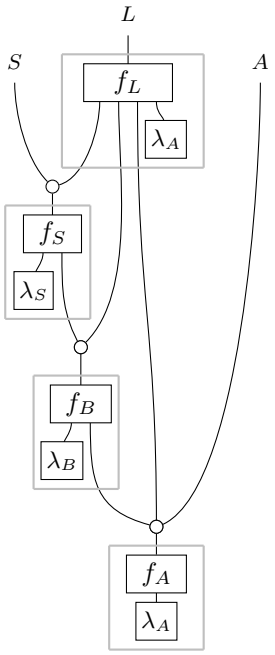


(26)

至关重要的是，将给定机制 c 分解为适当的 f、U、λ 的后一种方向具有高度的非唯一性。因此，存在大量包含变量 $V \cup U$ 的结构因果模型，它们都能产生具有变量 V 的同一个因果贝叶斯网络。这一事实本质上导致了因果层级第二层与第三层之间的分离。函数因果模型对于形式化反事实的重要性将在第8节中讨论。

例39。一个变量为 $V = \{B, S, L, A\}$ 和 $U = \{U_B, U_S, U_L, U_A\}$、且 V 之间的因果结构与例33相同的结构因果模型，由下方的网络图给出。对于 $X \in V$，可以选择（解释中的）f_X、λ_X 对，使公式（25）成立，其中左侧为相应的

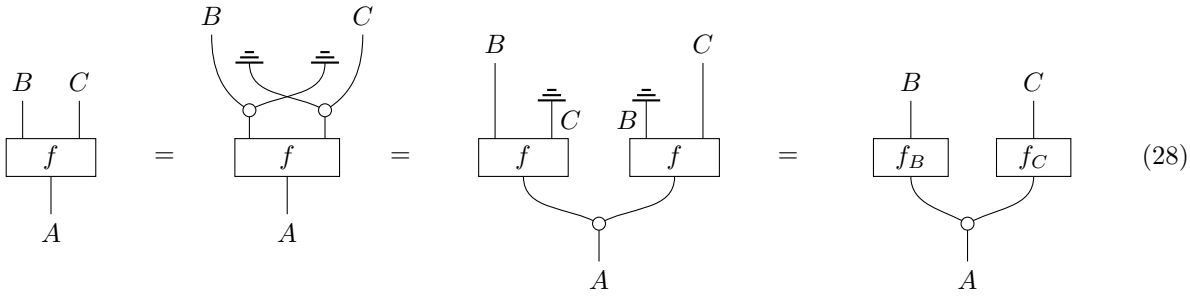
通道 道 c_X 来自例33中的因果贝叶斯网络示例，其中这些信道由灰色方框表示 s.



(27)

将因果关系视为最终由相对于某个真实函数（确定性态射）的依赖关系所定义（即使该函数未知），还有另一层重要意义：这为因果模型的定义中一项原本显得武断的规定提供了依据，即每个机制只有一个输出。关于这一论证的详情可参阅其他资料，例如参考文献 [4, 3, 83, 72, 84, 85]；我们指出，这一事实在函数因果模型的图示表示中直观地显现出来。

给定一个类似例39的函数因果模型，将这些对 f_X, λ_X 视为 $X \in V$ 的机制 c_X，并依据公式（25），我们得到一个定义21意义上的因果模型。然而，将机制断言为单输出通道 c_X: Pa(X) $\to X$ 与将其断言为单输出确定性信道 f_X: Pa(X) $\to X$ 之间的区别在于它们的因子分解性质。任何具有多个输出的函数，例如$f\colon A \to B \times C$，都可以按分量函数 f_B, f_C 进行因子分解，其输出为 B, C。更一般地，cd-范畴中的任何确定性态射都可以按这种方式因子分解：



(28)

对于两个以上的因子，同样的性质也可以迭代成立。[5] 关键的是，cd-范畴中的一般信道 $c\colon A \to B \otimes C$ 无法按这种方式进行因子分解，正如 FStoch 中的一般概率信道也无法如此因子分解。

因此，若仅将因果模型视为通过概率通道定义的模型，人们或许会倾向于允许通道具有多个输出；但若最终以函数依赖为因果关系的基础，便会得出这样的观点：机制应当只有单个输出。

[5]如果输入具有与不同变量相对应的乘积结构，那么可以将这一因子分解性质与重写确定性信道 f_B 和 f_C 结合起来，使其符合各自实际的依赖结构，从而让 f_B 和 f_C 的输入仅包含 B 和 C 分别通过 f_B 和 f_C 实际依赖的那些导线。

3.4 忠实性概念

因果模型框架的一个重要方面，是它在因果断言与条件独立关系之间建立了精确的联系。概率分布必须满足因果结构所蕴含的所有条件独立关系，才能与该因果结构相容。因此，这种联系也构成了因果发现算法解法的基础 [4, 5, 86] ——也就是这样一个认识论问题：给定一个概率分布，推断因果结构（所受的约束）。

这一联系通过 d-分离定理 [87, 88]，以纯图论的 d-分离概念形式化。给定一个有向无环图 G，对于顶点的不相交子集 Y、Z 和 W，若 Y 和 Z 被 W d-分离（附录 C.5 中的定义 111），则记作 $(Y \perp\!\!\!\perp Z|W)_G$。下文证明，按照我们的定义，d-分离通常是条件独立性的可靠判据；证明见附录 C.5。[6]

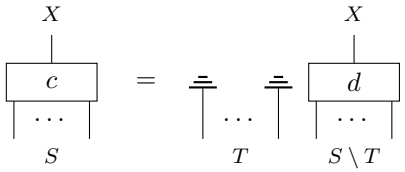
定理 40. 设 C 为一个具有效应条件化的 cd-范畴，且 ω 是 V 上的归一化态，它按照一个包含变量 V 的有向无环图 G 进行因子分解。那么，对于任意不相交子集 Y、Z、$W \subset V$，我们有

$$(Y \perp\!\!\!\perp Z|W)_G \quad \Rightarrow \quad (Y \perp\!\!\!\perp Z|W)_\omega \ .$$

反过来，若一个因果模型 $\mathbb{M}$ 在 C 中是忠实的，我们称其为忠实，当且仅当其诱导态 $\omega = [\![\mathbb{M}]\!]$ 满足的条件独立性仅限于因果结构所蕴含的那些条件独立性，即：$(Y \perp\!\!\!\perp Z|W)_\omega \implies (Y \perp\!\!\!\perp Z|W)_{G_\mathbb{M}}$。一般而言，忠实性概念的重要意义在于，它是因果发现中奥卡姆剃刀式的一种期望准则 [4]。

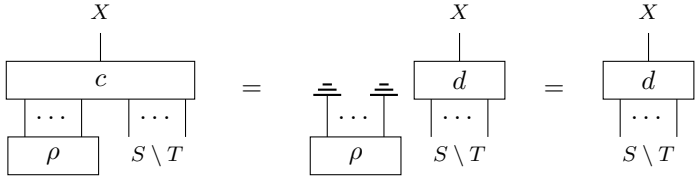
现在，由于一个实际的因果模型会为每个变量指定因果机制，因此我们还得到第二种在逻辑上彼此独立的忠实性概念。要说明这一概念，最简单的方式是先定义信道的以下性质。

定义 41。（无信号）：给定一个从输入 S 到 X 的信道 $c\colon S \to X$，若存在一个信道 d，使得



(29)

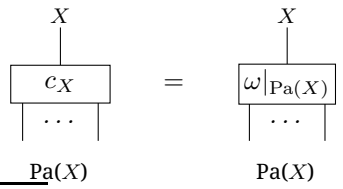
这一术语指的是，上述条件确保系统 T 的任意输入 ρ 都不会导致由 c 决定的 X 发生变化，也就是说：



(30)

给定一个具有变量 V 和有向无环图 $G_\mathbb{M}$ 的因果模型 $\mathbb{M}$，若机制 c_X: Pa(X) $\to X$ of $X \in V$ 不满足公式（29），则称其为忠实的；也就是说，不存在任何非空的 $T \subseteq$Pa(X) 使公式（29）成立。若因果模型 $\mathbb{M}$ 的所有机制都忠实，则称该模型机制忠实。因此，对每个变量 X，其每个父节点都有可能向 X 发送信号。以下结果有助于我们联系所讨论的忠实性概念。

引理 42. 设 $\mathbb{M}$ 是一个因果模型，其诱导态 $\omega = [\![\mathbb{M}]\!]$ 在某个变量 X 的 Pa(X) 上具有全支撑。那么



(31)

[6]最近，Fritz 和 Klingler 以纯粹的字符串图术语，提出了马尔可夫范畴中 d-分离的一种广义概念 [13]。然而，由于这项工作采用了不同的条件独立性概念（涉及部分信道），我们必须独立验证 d-分离的可靠性。

证明。附录 C.6。 □

由此结果可知，如果 $\mathbb{M}$ 是一个因果模型，且其诱导态在所有变量上都有全支撑，那么它的机制便可（通过条件化）唯一确定，而 $\mathbb{M}$ 的忠实性意味着机制忠实性。一般而言，公式（31）却不成立，而且不难看出，这两种忠实性概念互不蕴含。[7]

理解非忠实因果模型在概念上的重要性，一种方式是考察因果模型与函数因果模型之间的关系。假设我们有一个因果模型，其机制 c_X最终源自一个函数因果模型，并由函数依赖关系 f_X 和额外的"噪声"状态 λ_X 构成，如公式（25）所示。那么，完全可能实际上存在一条从 Y 经由 f_X 到 X 的因果影响（对于某个$Y \in$Pa(X)而言）；尽管对于 U_X 上的某个特定（经过精细调谐的）状态 λ_X，这种影响被"抵消"，从而产生了一个无效忠实性的有效机制 c_X。因此，我们必须允许存在机制不忠实的因果模型。

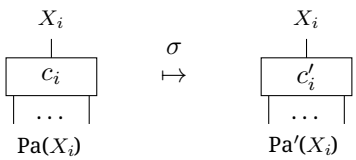
4 干预

4.1 定义

因果模型的一项主要优势在于，它能够描述如果对变量进行干预将会发生什么，即通过"介入"来修改因果机制，从而也总体上改变因果结构。大多数因果模型文献都聚焦于所谓的 do-干预，也称为原子干预——以至于"干预"通常就等同于"do-干预"。不过，诸如软干预或条件干预等更一般的干预类型也受到了关注，尤其是近年来。在此，我们首先在我们的语境中给出干预的一个非常一般的定义；这一定义与文献中因果贝叶斯网络和结构因果模型语境下的通常概念相一致（例如参见文献 [37]）。然后，我们将更常见的特定类型干预作为特殊情形重新得到。

定义 43. 设 C 为一个 cd-范畴。在具有变量 var($\mathbb{M}$) 的 C 中的因果模型 $\mathbb{M}$ 上进行的干预 σ，是对机制进行修改，从而得到一个在 C 中、具有相同变量 var($\mathbb{M}$) 的新因果模型 $\mathbb{M}'$。

我们通常将干预 σ 称为通过函数将 $\mathbb{M}$ 中的每个机制 c_i 映射到 $\mathbb{M}'$中对应机制 c_i' 的干预。



(32)

干预下，顶点、变量和输出变量保持不变，而有向无环图中的边可能发生变化，因此一般而言 $G_{\mathbb{M}} \neq G_{\mathbb{M}'}$ 也会改变。在第 5.3 节中，我们将讨论更一般的因果模型变换，这些变换可能改变模型的任何方面。

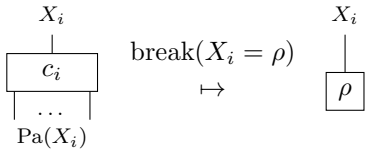
上述干预定义十分宽泛，涵盖了任意两个因果模型之间的转换，前提是变量集合保持不变。然而在实践中，干预通常只涉及一个或少数几个变量。因此，我们常用σ_X表示特定干预，其中 X 表示相对于原模型机制发生变化的变量子集；也就是说，隐含地认为，对于所有 $X_i \in V \setminus X$，机制均保持不变，$c_i' = c_i$。对于 FStoch 的情形，给定一个 CBN $\mathbb{M}$ 及其变量 V上的相应分布 $P(V)$，我们将干预 σ_X后因果模型 $\mathbb{M}'$所产生的分布记作 $P(V;\ \sigma_X)$。

[7]我们注意到，参考文献 [74] 引入了另一种条件独立性概念，称为强相对独立性。鉴于附录 F 中的分裂节点模型，参考文献 [74] 中从 $\mathrm{Mat}_{\mathbb{R}^+}$ 到 cd- 范畴中的因果模型，自然地推广了强相对独立性。这一概念在逻辑上强于（$Y \perp\!\!\!\perp Z|W$）$_\omega$，还根据 d- 分离关系引出了与之对应的忠实性概念。我们将三种忠实性概念之间关系的进一步探讨留待未来工作。

4.2 干预类型大全

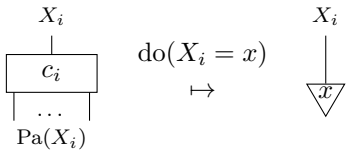
下面我们列出一些值得注意的干预类别，并举例说明它们在图示因果推理中的应用——当我们用网络图指定因果模型时，直观而言，干预相当于进行“弦图手术” [9]。

Br破坏干预。在 X_i 处的破坏干预是一种形式为



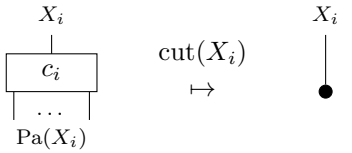
在修改后的有向无环图中， X_i 没有父节点；它被变成一个根节点，因为先前的机制 c_i 被覆盖，并由状态 ρ 取代。在 FStoch 中，这对应于为 X_i 指定一个概率分布，并构成「软干预」的一种特殊情形。例如，可以考虑「模糊 do 干预」：由于实际条件的限制，只能以一定概率将 X_i 固定为某个值。

事实上，当 ρ 为点分布时，就会出现 do-干预；更一般地，在任意 cd-范畴中，$\mathrm{do}(X_i = x)$ 是一种破坏干预，其中 $\rho = x$ 是 X_i 上的一个尖锐状态 x 。



(33)

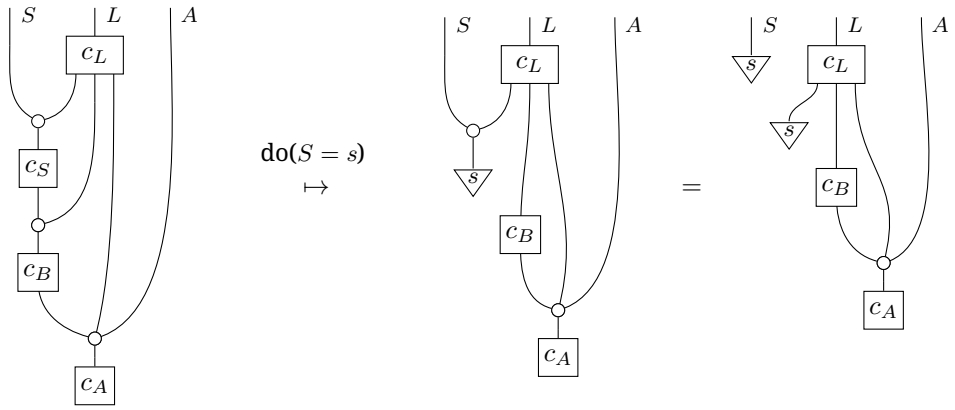
给定一个变量为 V 的因果贝叶斯网络，我们用 $P(V; do(X_i = x))$ 表示新模型对应的干预后分布。do-干预在因果效应可识别性问题中十分重要（见第7节）。作为破坏干预的另一种特殊情形，切断干预 $\mathrm{cut}(X_i)$ 是指 $\rho =$ 为 X_i [9] 上一个特定的归一化均匀状态。



(34)

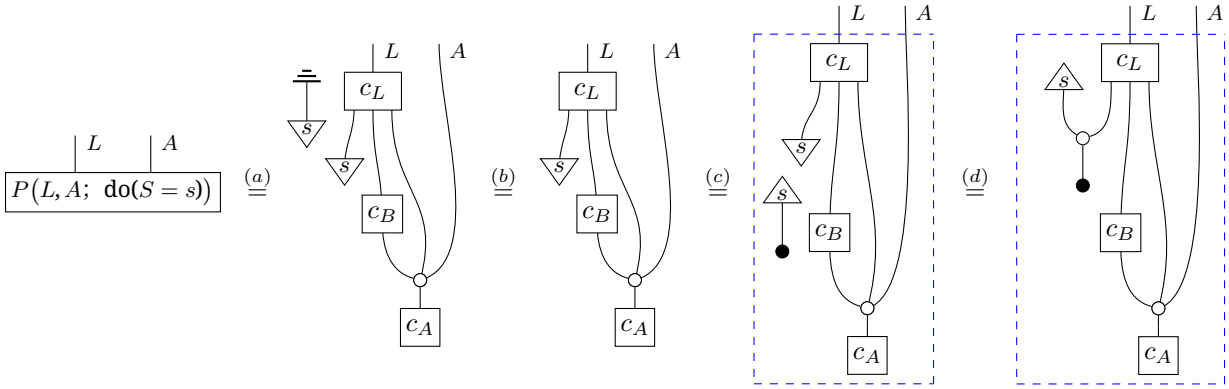
直观地说，切断干预会将变量的原始机制替换为其取值上的均匀随机分布。切断干预在 [9] 中占有重要地位。请参见后续示例，了解如何看待它们与 do-干预之间的关系。

例44。回到例33中的因果贝叶斯网络，考虑 do-干预 $\mathrm{do}(S = s)$，它强制 S 处于确定性状态 s。例如，这可能对应一项理想化的立法，强制所有人吸烟或不吸烟，具体取决于 s的值。该模型对输出分布$P(S, L, A)$的效应如下。



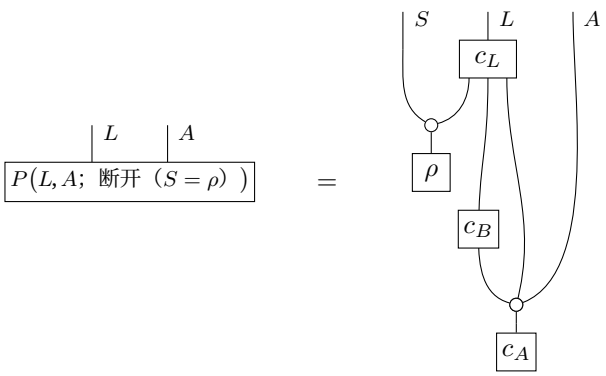
(35)

这里我们利用了确定性状态会“传递复制”的事实，参见式（4）。我们也用这个例子来说明 do($S = s$) 干预与 cut(S) 干预之间的关系，以及按照第 2 节的图示方式对$S = s$ 进行条件化：



上述等式体现了本著作中将频繁出现的一类图示改写：(a) 根据边缘化的定义成立，(b) 用到了状态 s 已归一化这一点，(c) 用到了式 (11) 以及状态具有全支撑这一点，而 (d) 用到了式 (6)。注意，(d) 的右侧不过是在经历了 cut(S) 的模型中对 $S = s$ 进行条件化；可以将其视为对随机对照试验的建模：在该实验中， S 接受的治疗（此处暂且不论这在伦理上有多大问题）是随机分配的，随后再对从该实验获得的分布中的 $S = s$ 进行条件化。

重要的是，对于一般的破坏干预，（不一定是确定性的）状态 ρ 不会像上面那样“通过”复制映射，从而得到乘积分布；S 与 L之间会存在非平凡的相关性：



(36)

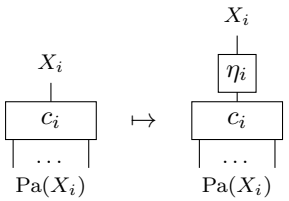
上述示例中的步骤（a）和（b）说明，对于因果贝叶斯网络$\mathbb{M}$的变量 V 的任意子集 $S \subseteq V$，do-干预的图示结果体现了针对 $P(V \setminus S; \text{do}(S = s))$的所谓截断因子分解公式。该公式要求从式（22）中的 $P(V)$中丢弃所有对应于 S的因子，并固定 $S_X = s_X$；只要 $X \in S$ 出现在某个不属于 S的变量的父变量中，就要这样做。

事实上，一般而言，这一分布是通过对相应的网络图 $D_{\mathbb{M}}$进行“手术”得到的：将每个 $X \in S$对应的机制 c_X 替换为确定性状态 s_X，然后让 s_X “传递复制”，并丢弃 $D_{\mathbb{M}}$ 中对应于 S的输出线。最后剩下的整体状态表示$P(V \setminus S; \text{do}(S = s))$，这确实与该公式一致。

图示的　　这种方法还便于超越破坏干预，例如以下　　g.

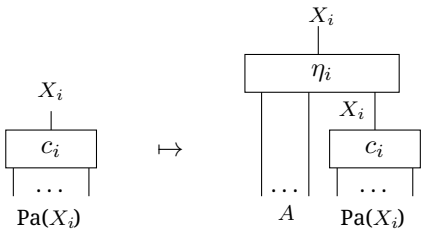
[8]这正是 Jacobs *et al.* 使用切断干预来建模 do-干预的含义，尽管参考文献 [11] 中的设置并不允许以图示方式处理条件化。

局部干预。在 X_i 处，局部干预是如下形式的干预



(37)

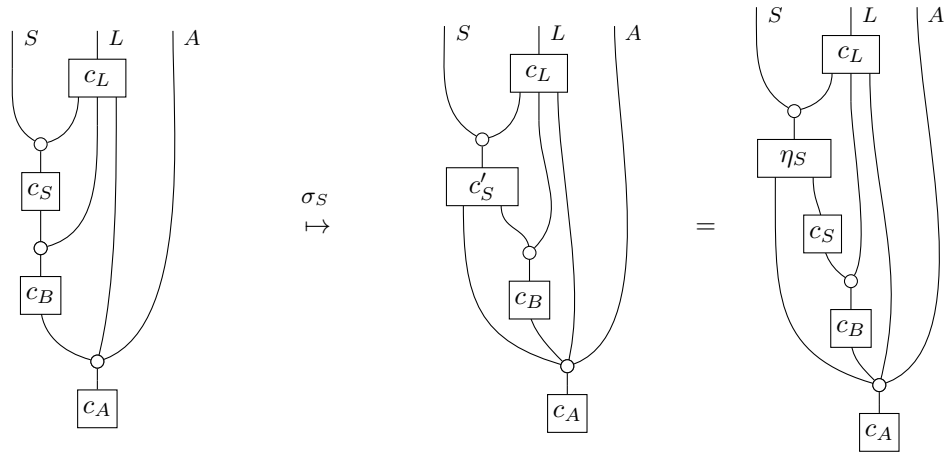
对某些 信道 η_i: $X_i \to X_i$。更一般地，在 X_i 处的宽局部干预形式为



(38)

对于某个信道 η_i: $A \otimes X_i \to X_i$，其中 $A \subset V \setminus Pa(X_i)$，“宽”表示引入新的直接原因，使其作用于 X_i。对于任一种局部干预，都可以设想有一个主体驻守在 X_i，在局部实施干预。信道 η_i 是对她依据机制 c_i接收到的系统所进行的“后处理”的有效描述。局部干预也与因果模型的所谓“分裂节点”表示中所考虑的那类干预密切相关 [74], , 见附录 F。

例 45. 再次回到贯穿全文的因果贝叶斯网络示例。例 44 考察了一个 do-干预 $\mathrm{do}(S = s)$，以及一种一般的破坏干预；后者是一种非原子干预，并且已经体现了软干预的一个方面。要改善公共卫生，一种更现实的政策或许是：实际上，所有 21 岁以下人群都有比如 90% 的概率不吸烟，而所有 21 岁以上的人则不受该政策影响 *[37]*。这可以通过一种宽局部干预来建模，该干预将 A 引入为 S 的新直接原因，形式如下：



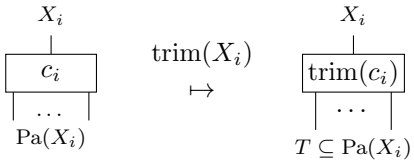
其中， S 的新机制可以用旧机制表示为：

$$c'_S(B, A) := \begin{cases} 0.9\ \delta_{S,0} + 0.1\ \delta_{S,1} & \forall A \leq 21 \\ c_S(B) & \forall A > 21 \end{cases} \tag{39}$$

公式 (39) 还允许 读出适当的信道 η_S，将其与 c_S 预先复合后即可得到 c'_S .

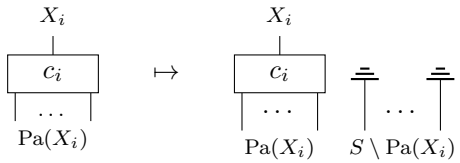
从更形式化的角度来看，我们可以将干预视为这样一个过程：从某个变量中移除某些“实际上无关”的因果父节点，或添加多余的因果父节点；我们分别将这两种操作称为“修剪”和“填充”。

修剪干预。 考虑一个机制 c_i 关于 X_i，并假设每个对象 $X_j \in \mathrm{Pa}(X_i)$ 至少有一个归一化态。在附录 D 中，我们证明，此时存在唯一的子集 $T \subseteq S = \mathrm{Pa}(X_i)$和态射 d，使得 c_i 可分解为 $_T \otimes d$，如公式（29）所示，并且 d 作为一个输入为 T的机制来看是忠实的。我们将 $d = \mathrm{trim}(c_i)$ 称为 的修剪，并将干预 $\mathrm{trim}(c_i)$ 定义为 $X_i\mathrm{trim}(c_i \mapsto)$。$c_i$。

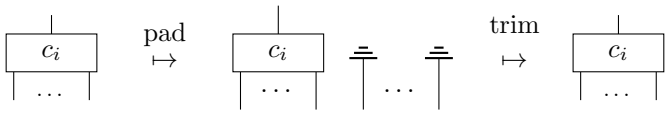


在有向无环图层面，这涉及移除所有边 $Y \to X_i$：对于这些边， Y 通过机制 c_i 实际上不会向X_i 传递信号（见定义 41）。当且仅当模型满足机制忠实性时，该模型在所有修剪下保持不变（见第 3.4 节）。

填充干预。反过来，对于任意变量子集 S 和变量 X_i，我们将填充干预 $\mathrm{pad}(X_i, S)$ 定义为

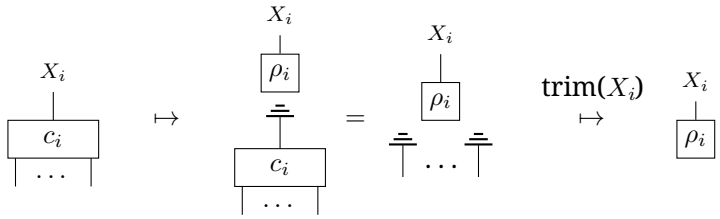


这在形式上将 S 中的变量设为 X_i 的父节点，并添加有向无环图中尚不存在的所有边 $S \to X_i$。根据构造，若 $S \not\subseteq \mathrm{Pa}(X_i)$，则 X_i 的新机制将不忠实。填充是修剪的逆操作：只要 c_i 忠实，且 S 与 $\mathrm{Pa}(X_i)$ 不相交，就有 $\mathrm{trim}(\mathrm{pad}(X_i, S))$ 保持 c_i 不变。



我们强调，修剪和填充本身既不是操作性的，其作用也不是淡化非忠实模型在概念上的重要性。后者对于断言存在直接原因关系 $X \to Y$ 至关重要，即使在改变X时无法对 Y产生任何影响，因为对 Y 的机制的有效描述恰好是某个特定的 c_Y。（参见第 3.4 节的讨论。）尽管如此，修剪和填充仍是便于表述一些显而易见但有用事实的形式体系工具，包括以下内容。

注释 46。（将破坏干预视为丢弃和修剪）：破坏干预 $c_i \to \rho_i$ 可以看作标准局部干预的复合，即将 c_i 与 $\rho_i \circ$ 复合，随后进行修剪干预，如下所示：

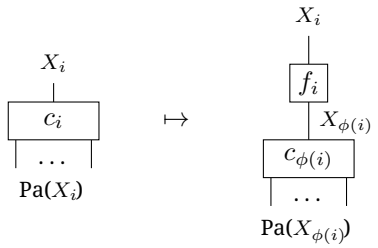


正是在这个意义上，引言中图 2 的示例表示了一种 do-干预。在后文各节讨论示例时，我们也经常会以这种方式描绘 do-干预：丢弃原有机制，随后进行简化重写，其中尤其包括修剪。

在因果模型的框架中，干预能够改变一个机制而不影响任何其他机制，这仅仅是形式体系定义使然。假设在实践中确实可以

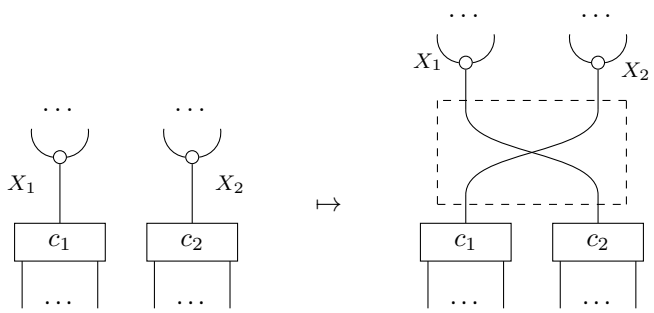
做到这一点，通常称为机制的自主性。[9] 迄今为止我们遇到的特定类别的干预可以描述为“乘积干预”，因为每个变量所对应的机制都不依赖于其他变量的机制。为了说明干预不一定要采用“乘积形式”，同时机制仍保持自主这一性质，请看以下最后一类干预。

重连干预。给定用于 $X_1, \dots, X_n$的机制 $c_1, \dots, c_n$、 $\{1, \dots, n\}$ 的一个置换 ϕ 以及态射 f_i: $X_{\phi(i)} \to X_i$，我们可以定义一种重连干预:



用于 $i = 1, \dots, n$。在有向无环图层面，这相当于按照ϕ置换顶点的标签，同时保持边“原样不动”。更准确地说，新模型中的父节点集为 Pa′(X_i) =Pa($X_{\phi(i)}$)。在机制层面，变量 X_i 现在由用于 $X_{\phi(i)}$的机制给出，并经过变换 f_i。这种干预在网络图层面最为自然，因为它相当于“交换导线”，然后与 f_i复合。

例47。考虑同一对象上有两个变量 $X_1 = X_2$ 的情况，并考虑一种重连方式：交换这两个变量，同时保持其他变量不变，其中 $f_1 = f_2 = \mathrm{id}_{X_1} = \mathrm{id}_{X_2}$。这相当于在网络图中插入一个交换，如下所示:



因此现在 X_1 具有机制 c_2，而 X_2 具有机制 c_1。

关于所列干预类型如何在具体因果模型中体现为“图式手术”，更多示例可见第7节。

5 开放因果模型

5.1 定义

在定义干预时，我们看到，干预往往实际上只作用于机制的一个子集，而非整个因果模型。形式上，这样的机制子集可以视为一个“带输入的因果模型”，同时也带有输出，因为它可能包含一些未指定其机制的变量，而这些变量可以解释为模型的输入。

[9]这也可以表述为一种假设：人们能够以恰当的方式选择因果关系项，使其机制确实是自主的，并使因果模型形式体系得以适用。理解这一假设何时适用，以及它适用的确切含义，是当前研究的主题；例如参见文献 [89, 68]

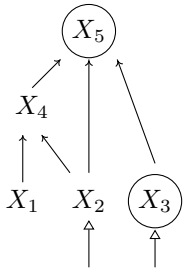
本节将第3节中的处理扩展到这类“开放”因果模型。尽管对它们的研究不那么广泛，但从范畴论视角来看，它们实际上非常自然；尤其是在与第3.2节中的网络图相对应时、处理（打破）干预时，以及我们在第8节讨论反事实时更是如此。我们首先给出一个合适的带有“输入和输出”的有向无环图概念。

定义 48. 开放有向无环图 (G, I, O) 由一个有限有向无环图 G 给出，该图的顶点为 V，并配有一个输出子集 $O \subseteq V$ 和一个输入子集 $I \subseteq V$，使得每个输入 $i \in I$ 在 G 中都没有父节点。

注意，我们不要求 I 和 O 互不相交，因此一个顶点可以同时是输入和输出。为简便起见，我们通常直接用 G 来指代这样的开放有向无环图。

开放有向无环图 (G, I, O) 的图示方式与有向无环图 G 相同，只需用圆圈突出显示输出变量 O，并为每个输入顶点 $i \in I$ 画一条没有源点的特殊输入箭头指向该顶点。关于开放有向无环图与参考文献 [39, 40] 中的“开放图”有何关联，见注释 59。

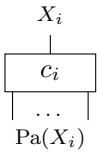
例49. 下图展示了一个定义在 $V = \{X_1, \ldots, X_5\}$ 上的开放有向无环图，其输入为 $I = \{X_2, X_3\}$，输出为 $O = \{X_3, X_5\}$。



开放有向无环图 s 指定了我们相应概念的因果结构，即“带输入的因果模型 ’.

定义 50. 设 C 是一个 cd-范畴。C 中的一个开放因果模型 $\mathbb{M}$ 由以下部分给出：

1. 一个开放有向无环图 $G = (G, I, O)$，其顶点为 V；2. 对每个顶点 $X_i \in V$，关联一个 C 中的对象，也记作 X_i；3. 对每个顶点 $X_i \in V \setminus I$，关联一个 C 中形式如下的信道：



(40)

称为 X_i 的机制。

开放有向无环图 $G = (G, I, O)$ 称为 $\mathbb{M}$ 的因果结构，也记作 $G_{\mathbb{M}}$。C 中由 V 索引的对象称为 $\mathbb{M}$ 的变量 var($\mathbb{M}$)。由 I 和 O 索引的对象分别称为输入变量和输出变量，并记作 in($\mathbb{M}$)、out($\mathbb{M}$) $\subseteq$ var($\mathbb{M}$)。若开放因果模型 $\mathbb{M}$ 没有输入，即 in($\mathbb{M}$) $= \emptyset$，则称其为闭合的；这也就是说，它构成了我们先前意义上的因果模型。

换句话说，具有变量 var($\mathbb{M}$) 的开放因果模型 $\mathbb{M}$ 的定义与具有相同变量 var($\mathbb{M}$) 的（封闭）因果模型相同，唯一的区别是没有为输入变量 in($\mathbb{M}$) 指定机制。我们将其余（非输入）变量称为被致因变量，将非输出变量称为内部变量，并记作如下形式。

$$\begin{aligned} \mathrm{caus}(\mathbb{M}) &:= \mathrm{var}(\mathbb{M}) \setminus \mathrm{in}(\mathbb{M}) \\ \mathrm{int}(\mathbb{M}) &:= \mathrm{var}(\mathbb{M}) \setminus \mathrm{out}(\mathbb{M}) \end{aligned}$$

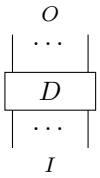
We 将模型 $\mathbb{M}$ 的机制（40）集合记为 mech($\mathbb{M}$) :=$(c_i)_{X_i \in 因果(\mathbb{M})}$。

与之前一样，我们可以定义一个开放因果模型 $\mathbb{M}$，而不必明确指定因果结构 G，只需给出一个对象的指标化集合 var($\mathbb{M}$)，及其子集 in($\mathbb{M}$) 和 out($\mathbb{M}$)，并为每个非输入变量 X_i 指定一个信道（见式（40））。要使模型有效，所诱导的有向图 G必须是无环的，其中 G 的每个变量对应一个顶点，并且当 $X_j \notin$ 属于 in($\mathbb{M}$) 且 X_i 是机制 c_j 的输入时，图中有一条边 $X_i \to X_j$。然后，我们定义相应的诱导开放有向无环图 $G_{\mathbb{M}} = (G, \text{in}(\mathbb{M}), \text{out}(\mathbb{M}))$。

我们关于因果模型与网络图之间的对应关系，现在可以直接扩展到包含输入的情形，如下所示。

Prop命题 51. 在等价关系及图示的重新标记意义下，以下对象等价 t:

- 一个定义在顶点 V 上的开放有向无环图 $G = (G, I, O)$;
- 一个网络图

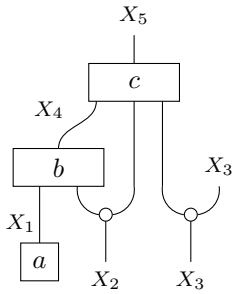


其线缆标签为 V，输入为 I，输出为 O。

证明。给定一个开放有向无环图 $G = (G, I, O)$，我们按照定义 25 中的方法定义网络图 D_G，但每个输入 $i \in I$ 不分配机制，而是直接作为图的输入，并复制到其子节点各自的机制中；此外，当 $i \in O$ 时，它也作为输出。

反过来，给定一个具有输入 I 和输出 O 的网络图 D，我们定义开放有向无环图 $G_D = (G, I, O)$，方法与命题 29 中定义 G_D 和 O 相同，只是还要为通向 D 的每条输入线添加一个输入顶点 $i \in I$。与之前一样，很容易看出，这些对应关系在重新标记和图的等价意义下互为逆映射。 □

例52。例49中的开放有向无环图对应于以下网络图。



一个开放因果模 型具有这种形式，意味着要像上文那样指定对象 X_1、.....、X_5 以及信道 a、b、c e.

由此可立即得到开放因果模型的图示刻画，与之前一样，不过现在允许网络图带有输入。

定理53。设 C 为一个 cd-范畴。在 C 中指定一个开放因果模型 $\mathbb{M}$，等价于指定一个带有 C 中解释 $[\![-]\!]$ 的网络图 D。

因此，简而言之，我们可以说，C 中的开放因果模型 $\mathbb{M}$ 是 C 中的网络图。对于任意开放模型，我们将 $[\![\mathbb{M}]\!] := [\![D_{\mathbb{M}}]\!]$ 记作 C 中从其输入到输出所得到的信道。

注54. 给定一个由状态网络图 D 及其解释$[\![-]\!]$所表示、属于 C 的（封闭）因果模型 $\mathbb{M}$，限制到子网络图 D'（此时带有输入）后，通过 $[\![-]\!]$再次得到一个开放因果模型。在 *[7],* 中，Fong 将因果变量之间的这种诱导态射 $[\![D']\!]$ 称为因果条件。

与开放模型相关的一个主要直觉和应用场景，来自处理第 4 节中的干预。可以把开放模型的输入线 X 看作对应于“do-干预 $\mathsf{do}(X = x)$，直至指定状态 x”，或者实际上更一般地，对应于“破坏干预 $\mathsf{break}(X = \rho)$，直至指定 ρ”。

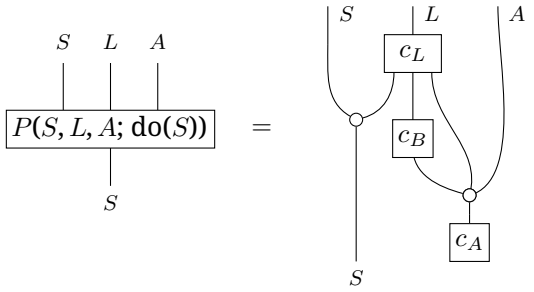
更准确地说，给定一个具有变量 V 和输出 $O \subseteq V$ 的（闭合）因果模型 $\mathbb{M} \in \mathsf{C}$，我们可以通过移除 X 的机制，定义一个具有变量 V、输入 X 和输出 O 的开放模型 $\mathbb{M}'$。表示 $Y \subseteq O$ 的常见表达式 $P(Y; \mathsf{do}(X))$，恰好就是由这个开放模型 $\mathbb{M}'$ 所诱导的总体通道 $[\![\mathbb{M}']\!]$（在 Y 上的边缘分布）；将输入状态 x 输入该通道后，会得到分布 $P(Y; \mathsf{do}(X = x))$，这是对 $\mathbb{M}$ 进行 do-干预 $\mathsf{do}(X = x)$ 所产生的结果。



(41)

在讨论 do-干预时，我们经常会在这个意义上使用记号 $P(Y; \mathsf{do}(X))$。为了具体说明，请看下面的例子。

例 55. 回到例 44，我们现在可以写成：

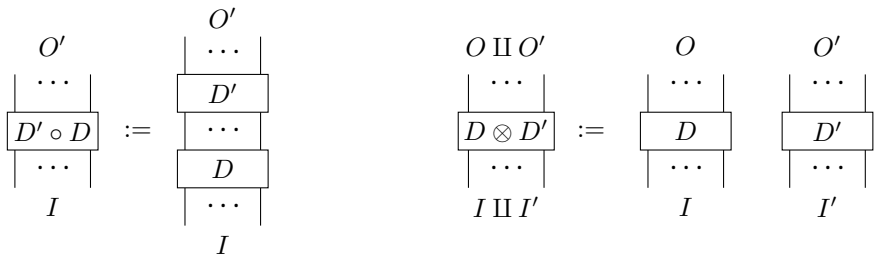


也就是说，对于右侧给出的网络图所对应的开放因果模型，有 $P(S, L, A; \mathsf{do}(S)) = [\![\mathbb{M}]\!]$。代入确定性状态 s 或一般状态 ρ，分别可得式 (35) 和式 (36)。

5.2 组合开放模型

允许同时存在输入和输出后，开放模型、网络图和开放有向无环图便都可以进行复合；事实上，它们各自都构成一个范畴。

网络图的复合或许最容易理解。给定网络图 D, D'，我们可以通过将它们并排绘制来进行并行复合 $D \otimes D'$；也可以通过将它们连接起来进行顺序复合 $D' \circ D$，前提是 D 的输出 O 与 D' 的输入 I 相匹配。



(42)

其中 $S \sqcup T$ 表示集合 S 和 T 的不交并。形式化地，我们有如下定义：所谓指标化集合 $(X_i)_{i \in I} \simeq (Y_j)_{j \in J}$ 的重新标记，是指一个双射 $f\colon I \to J$，满足对所有 $i \in I$，都有 $X_i = Y_{f(i)}$。

命题56。网络图构成一个对称幺半范畴 ND，其对象是有限集，而态射 $D\colon I \to O$ 是配备了输入 $I_D \simeq I$ 和输出 $O_D \simeq O$ 重新标记的网络图。

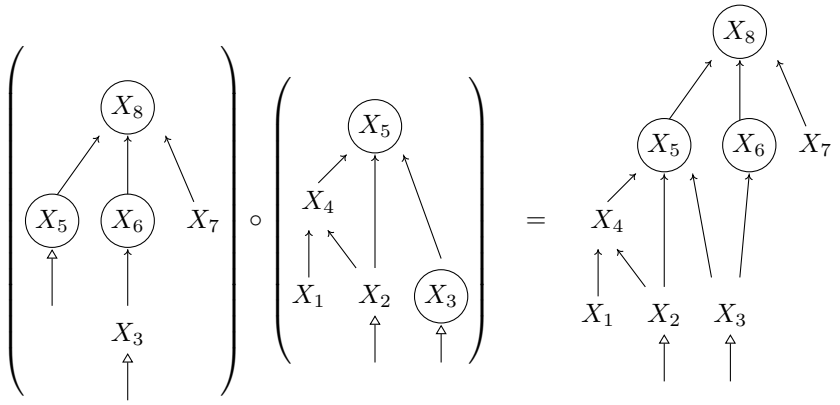
证明。我们按照式 (42) 定义复合，其中 I 上的恒等图完全由恒等线组成，并且 $X \otimes Y = X \amalg Y$ 定义在对象上，且 $I = \emptyset$。具体而言，设 L_D 表示图 D 的线缆标签。那么 $L_{D \otimes D'} = L_D \amalg L_{D'}$ 由不交并给出，而 $L_{D' \circ D} = L_D \amalg L_{D'} / \sim$ 则是将 I' 与 O 识别后得到的商。 □

根据命题 51 中网络图与开放有向无环图之间的对应关系，我们立即得到一个与之同构的开放有向无环图范畴。

推论 57. 开放有向无环图构成一个对称幺半范畴 ODAG，它与 ND 同构；其对象是有限集，态射 $G\colon I \to O$ 是带有重新标记 $I_G \simeq I$, $O_G \simeq O$ 的开放有向无环图。

证明. 同样，对象 $X \otimes Y = X \amalg Y$、$I = \emptyset$ 和 $G \otimes G'$ 分别由输入、输出和有向无环图的不交并给出。具体而言，对于 $G\colon I \to O$ 和 $G'\colon I' = O \to O'$，$G' \circ G$ 仍然是对不交并取商，方法是识别 $O = I'$ 中对应的元素。X 上的恒等态射是具有 $V = I = O = X$ 且没有边的开放有向无环图。 □

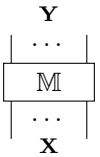
例 58. 使用前面引入的记号，下面给出了开放有向无环图复合的一个例子。



注 59. 在 *[39]*, 中隐含地、在 *[40]*, 中明确地，开放图 $G\colon I \to O$ 被定义为一个有向图 $G = (V, E)$，并带有函数“输入” $i\colon I \to V$和“输出” $o\colon O \to V$。开放有向无环图则是满足 i 和 o 均为单射的有限开放图；将它们视为子集 I、$O \subseteq V$，并且每个 $i \in I$ 在 G 中都没有父节点。开放图构成一个对称幺半范畴 OGraph，其复合同样通过识别相应的输出和输入来定义（形式上通过推出）。因此，ODAG 是 OGraph 的一个幺半子范畴。

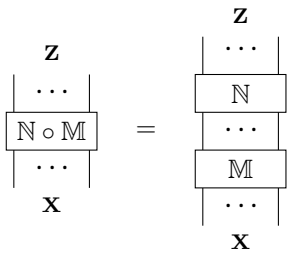
由于开放因果模型只是附带一种解释的网络图，因此我们可以像复合网络图那样复合开放模型；它们也构成自己的范畴。

推论 60. C 中的开放因果模型构成一个对称幺半范畴 OCM(C)，其对象是 C 中对象的有限指标化集合 $\mathbf{X} = (X_i)_{i=1}^n$，而态射

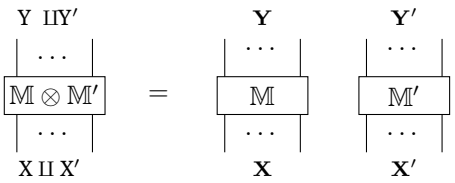


是 C 中的开放因果模型 $\mathbb{M}$，配备了重新标记 $\mathsf{in}(\mathbb{M}) = \mathbf{X}$ 和 $\mathsf{out}(\mathbb{M}) = \mathbf{Y}$。这里 $\mathbf{X} \otimes \mathbf{Y} = \mathbf{X} \amalg \mathbf{Y}$ 且 $\mathbf{I} = \emptyset$。

具体而言，给定模型 $\mathbb{M}\colon \mathbf{X} \to \mathbf{Y}$ 和 $\mathbb{N}\colon \mathbf{Y} \to \mathbf{Z}$，它们的顺序复合：

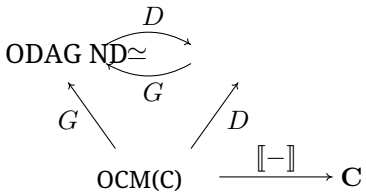


由组合它们对应的网络图给出，使得 $D_{\mathbb{N}\circ\mathbb{M}} = D_{\mathbb{N}} \circ D_{\mathbb{M}}$，并取来自每个模型的机制的（不交）并集 $\mathbf{mech}(\mathbb{M}) \amalg \mathbf{mech}(\mathbb{N})$。变量 $\mathbf{var}(\mathbb{N} \circ \mathbb{M})$ 构成 $\mathbb{M}$ and $\mathbb{N}$ 的变量的不交并的商，其等价关系确定了它们在 $\mathbf{Y}$ 中相匹配的位置。类似地，给定模型 $\mathbb{M}\colon \mathbf{X} \to \mathbf{Y}$ 和 $\mathbb{M}'\colon \mathbf{X}' \to \mathbf{Y}'$，它们的并行复合



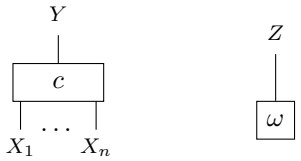
通过将图示并排绘制来给出，其中 $D_{\mathbb{M}\otimes\mathbb{M}'} = D_{\mathbb{M}} \otimes D_{\mathbb{M}'}$，并取各模型中变量与机制的不交并。

注 61. 我们可以用以下幺半函子的交换图概括本节中的范畴，其中 $\simeq$表示范畴同构：

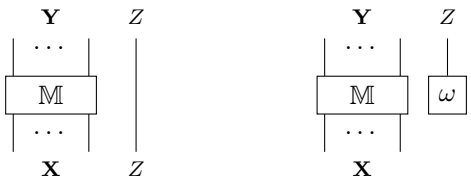


这里 $[\![-]\!]$ 是由 $\mathbf{X} \mapsto \bigotimes_{i=1}^{n} X_i$ 和 $\mathbb{M} \mapsto [\![\mathbb{M}]\!] := [\![D_{\mathbb{M}}]\!]$给出的函子，它将一个开放因果模型映射到 $\mathbf{C}$ 中从其输入到输出的单个信道。

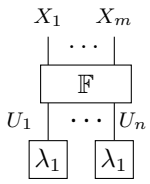
开放模型的组合提供了一种从已有因果模型构建新因果模型的方法；下面是几个基本示例。首先注意，$\mathbf{C}$ 中的任意单个信道 $c\colon X_1 \otimes \cdots \otimes X_n \to Y$都构成一个开放因果模型 $c\colon \mathbf{X} = (X_i)_{i=1}^n \to (Y)$，其中 c 是其唯一的机制。特别地，$\mathbf{C}$ 中 Z 的任意归一化态 ω定义了一个闭合因果模型 $\omega\colon \emptyset \to (Z)$。因此，我们可以将下列对象视为 $\mathbf{C}$ 或 $\mathbf{OCM(C)}$ 中的态射。



例 62. 作为并行组合的一个简单例子，考虑在因果模型 $\mathbb{M}$ 中添加一个没有父节点的额外输出 Z。左侧展示的是不给它分配 Z 任何机制、使其成为输入的情况，而右侧展示的是给它分配由归一化态 ω 给出的机制的情况。



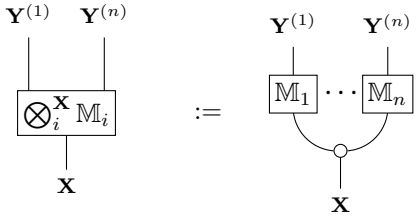
例63. 如果一个开放因果模型只有确定性机制，我们就称它为确定性的。现在，函数因果模型（定义3.3）也可以另行定义为以下形式的（封闭）因果模型的边缘分布：



对于某些归一化态 λ_i，其属于 U_i，对于 $i = 1$、.....、n，其中 $\mathbb{F}$ 是一个确定性模型，其变量为X_1、.....、X_n、U_1、.....、U_n，并满足每个 X_i 都只有一个输入 U_i 作为父节点，且每个 U_i 都有一个子节点。这里 $m \leq n$，其中 X_1、.....、X_m 是输出变量。

除了顺序复合和并行组合之外，还有最后一个概念：虽然它并非 OCM(C) 中模型的复合，却仍能将若干开放模型组合成一个新模型。我们称之为变量共享。在第8.1.2节处理反事实时，构造“平行世界”模型特别适合采用这种方法。

定义64。给定具有相同输入 $\mathbb{M}_i$ 的开放因果模型：$\mathbf{X} \to \mathbf{Y}^{(i)}$，对于 $i = 1$、.....、n，我们可以构造一个新的开放模型，记作 $\mathbb{M} := \bigotimes_i^{\mathbf{X}} \mathbb{M}$。该模型共享这些输入；最自然的定义方式是通过其网络图，即将各模型的网络图复合起来，并将输入变量 X 复制后传递给每个模型：



请注意，上方右侧的图本身并不是 OCM(C) 中的图，因为后者不包含复制。

5.3 因果模型的变换

我们已经看到，可以通过按照它们所构成的对称幺半范畴中的方式组合因果模型，以及共享变量，来构建新的因果模型。本节将介绍开放因果模型和开放有向无环图的若干其他变换，这些变换使我们能够构造新的模型和图。每种变换既可作为对开放模型的操作，也可作为对开放有向无环图的操作；我们将分别说明二者。这些变换涵盖了人们在实践中通常会对因果模型进行或利用因果模型进行的许多“操作”，从而使它们以明确的方式成为形式体系的一部分。

5.3.1 内化与外化

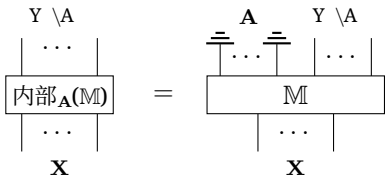
第一种变换是简单地改变哪些变量被视为输出，从而变换模型。第 3.2 节已经提到，在网络图层面，丢弃和复制输出互为逆元。它们在模型层面上的表现对应于 OCM(C) 中的变换。

首先需要指出，虽然 OCM(C) 缺少复制，但它具有丢弃效应。对于任意变量集合 X，丢弃效应：



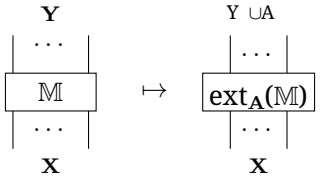
是一个输入为 X、没有输出且没有机制的开放因果模型。

现在，给定一个因果模型 $\mathbb{M}$ 及其输出的子集 A，我们可以构造一个新的因果模型 $\text{int}_{\mathbf{A}}(\mathbb{M})$，称之为将 A 内部化，具体如下。[10]



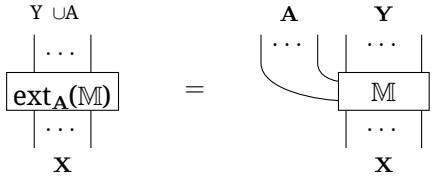
模型 $\text{int}_{\mathbf{A}}(\mathbb{M})$ 与 $\mathbb{M}$ 具有相同的变量、输入和机制，但变量 A 不再被视为输出。这相当于替换开放有向无环图 $G = (G, I, O)$ 的 $\mathbb{M}$ 由 $G = (G, I, O \setminus A)$。请注意，OCM(C)的结构不止对称幺半范畴的结构。特别是，我们可以“深入”态射内部执行额外操作。

给定一个因果模型 $\mathbb{M}$ 和子集 A $\subseteq$ var($\mathbb{M}$)，我们定义 A 的外部化：

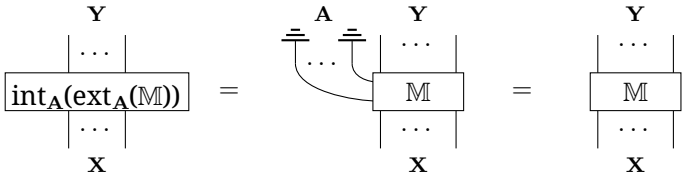


其输入和机制与 $\mathbb{M}$ 相同，输出则由 out($\mathbb{M}$) ∪A 给出。在开放有向无环图层面，我们将 $G = (G, I, O)$ 替换为 $(G, I, O \cup A)$。W

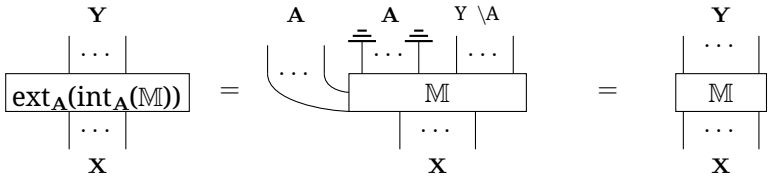
每当 A 与 Y =out($\mathbb{M}$) 不相交时，将外化表示如下是有意义的 s.



请注意，外化与内化互为逆运算，也就是说，每当 A 与Y 不相交时，我们有：



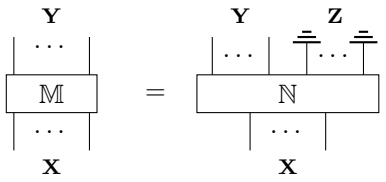
并且每当 A ⊆Y 时，我们有：



按定义 tion 两种运算都是幂等的：$\text{ext}_{\mathbf{A}}(\text{ext}_{\mathbf{A}}(\mathbb{M}))$ =$\text{ext}_{\mathbf{A}}(\mathbb{M})$，$\text{int}_{\mathbf{A}}$ 也同样如此 .

[10]我们再次强调，这个图并不是 C 中的网络图，而是 OCM(C) 中的图；其中的丢弃操作在形式上不同于 C 中的丢弃操作，但其效应与给定 C 中的网络图丢弃对应于 A 的线相同。

注 65. OCM(C) 具有以下“扩张”性质。我们称 $\mathbb{M}$ 为极大，如果每当 er



(43)

那么事实上 $\mathbf{Z} = \emptyset$ and $\mathbb{N} = \mathbb{M}$。等价地，$\mathbb{M}$ 是极大的，当且仅当 $\text{out}(\mathbb{M}) = \text{var}(\mathbb{M})$。于是，每个 $\mathbb{M}$ 都具有式 (43) 的形式，且对应唯一一个极大 $\mathbb{N}$，即 $\mathbb{N} = \text{ext}^{\text{var}(\mathbb{M})}(\mathbb{M})$，它是通过将 $\mathbb{M}$ 的每个变量都设为输出而得到的。

注 66。注意，式 (43) 仅当 $\mathbb{M}$ 和 $\mathbb{N}$ 具有完全相同的变量 $\text{var}(\mathbb{M}) = \text{var}(\mathbb{N})$、输入和机制，且仅在将哪些变量视为输出这一点上有所不同时才成立，其中 $\text{out}(\mathbb{M}) \subseteq \text{out}(\mathbb{N})$。因此，除了一个态射（模型）的输入和输出之外，通常还有许多变量存在于“内部”。特别是，OCM(C) 有许多标量：

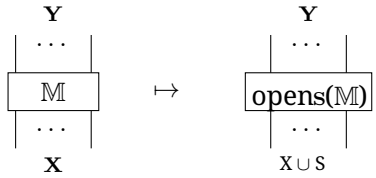


因为这些标量与没有输出的闭合因果模型一一对应。

5.3.2 开放与干预

迄今为止所考虑的变换，要么组合模型，要么仅仅调整哪些变量被视为输出。接下来介绍的变换则会改变机制，更具体地说，是移除其中一些机制。

给定一个从 $\mathbf{X}$ 到 $\mathbf{Y}$ 的模型 $\mathbb{M}$ 和子集 $\mathbf{S} \subseteq \text{var}(\mathbb{M})$，我们定义开放变换



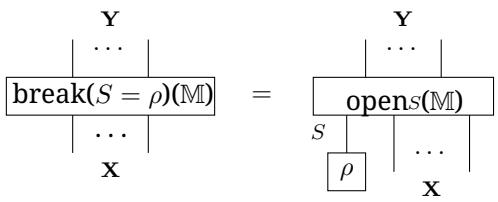
其中，模型 $\text{opens}(\mathbb{M})$ 是通过将每个 $S \in \mathbf{S}$ 设为输入而得到的，因此会从模型中移除其机制（如果存在）。因此，$\text{opens}(\mathbb{M})$ 的机制集为 $\{c_X \in \text{mech}(\mathbb{M}) \mid X \in \text{caus}(\mathbb{M}) \setminus \mathbf{S}\}$，其输入为 $\text{in}(\mathbb{M}) \cup \mathbf{S}$，并且具有相同的变量和输出集。

这对应于将开放有向无环图 $G = (G, I, O)$ 替换为开放有向无环图 $(G_{\overline{S}}, I \cup S, O)$，其中 $G_{\overline{S}}$ 是从 G 中移除所有指向 S 中顶点的边后得到的有向无环图。因此，在所得的开放有向无环图中，S 中的每个顶点都没有父节点，并被标记为输入。

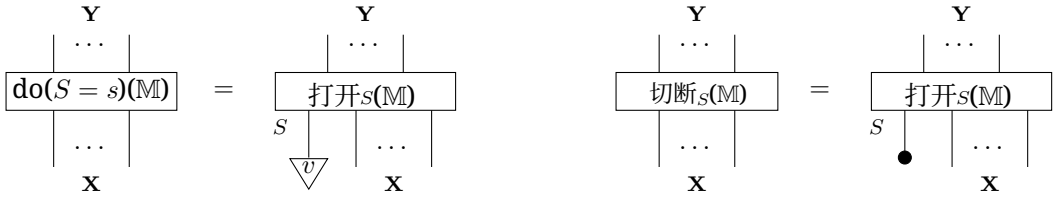
注释 67。开放与针对所谓 CADMGs 定义的固定操作密切相关；在该操作中，变量 S 的机制会被移除，随后通常会被设为某个预定义值 s。对于此处考虑的模型，固定等价于执行 do-干预 $\text{do}(S = s)(\mathbb{M})$。

开放因果模型与 do-干预，以及更一般的破坏干预之间的关系，已在第 5.1 节末尾提及。现在，我们可以用一种让干预本身成为模型变换的方式来描述这种关系，这种方式称为开放。

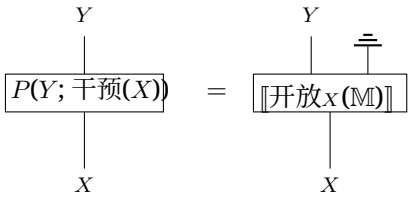
给定一个开放模型 $\mathbb{M}$，我们可以像对封闭模型那样，为 $S \subseteq \text{caus}(\mathbb{M})$ 定义一个破坏干预 $\text{break}(S = \rho)$。它由以下开放模型给出：



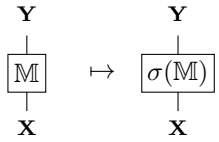
因此，破坏 S 的机制并将其替换为 ρ，相当于在 S 处开放模型 $\mathbb{M}$，使其成为一个输入，然后将它与 ρ 复合。$\text{do}(S = s)$ 和 $\text{cut}(S)$ 的特殊情形如下：



do-干预 $P(Y; \text{do}(X))$ 的总体通道在指定状态之前如下所示（因此，在第 5.1 节的讨论中，我们在公式（41）中有 $\mathbb{M}' = \text{open}_X(\mathbb{M})$）。



在上述对破坏干预的论述中，我们提到了开放模型上的干预这一一般概念。与闭合因果模型类似，我们将对开放因果模型 $\mathbb{M}$ 的干预定义为：将其重新指派为一个新的开放因果模型 $\sigma(\mathbb{M})$，该模型具有相同的变量 $\text{var}(\sigma(\mathbb{M})) = \text{var}(\mathbb{M})$、输入和输出：



因此，$\sigma(\mathbb{M})$ 仅在机制的选择上有所不同，因此其底层开放有向无环图也不同。我们在第4节讨论过的所有干预类别，同样都可以按照相同方式视为对开放因果模型的干预。同样，干预相当于对描述开放因果模型的网络图进行“手术”。

除此之外，开放因果模型的变换还有许多其他自然变体。有些变体可以直接在范畴 OCM(C) 的层面上表述，例如模型的复合和内化；另一些变体，如外化和一般干预，则需要“深入”态射内部，考察其中的底层因果模型。能够这样做，体现了开放因果模型远不止是一个范畴。至于它们究竟还满足哪些形式结构，我们留待未来研究。

6 含潜变量因果模型的处理

在实践中，一种常见情形如下。我们有一组观测变量O 的观察数据，因此可以估计相应的概率分布 $P(O)$；同时，我们对 O中的变量之间的直接因果关系掌握了一些但不完整的知识。这种知识之所以不完整，是因为我们知道存在潜在（未观测到的）混杂变量 L，或者无法排除其存在的可能性；该变量是 O中某些变量的共同原因。简言之，我们掌握了 $P(O)$，但不能假设 O 是由因果充分变量构成的集合。

因此，我们需要一种表示因果知识的方法，既能编码哪些地方可能存在或不存在潜在共同原因，也能说明这类部分知识如何约束观测分布 $P(O)$。关于各种能够处理潜变量的图模型，已有大量文献（例如，参见参考文献 [4, 90, 38]）。

我们在定义21中对（封闭）因果模型的定义，涉及所有变量 V 的因果机制，以及输出变量的一个子集 $O \subseteq V$。尽管“输出”一词是一个形式化且刻意保持中性的概念，

子集 O 通常确实对应于因果建模中通常意义上的“观测到的”或可访问变量。本节将说明，已有研究中的潜变量模型与我们对因果模型的图示定义之间有何关联。

6.1 潜在投影与无向混合图

以下是一个有用的图论概念。

定义 68.（无环有向混合图（祖先有向混合图）[11]）：无环有向混合图（祖先有向混合图） G 是一个元组（V, E, B），其中顶点集合为 V，有向边为 $E \subseteq V \times V$，且（V, E）是一个有向无环图，而 B 是顶点的无序对集合（即 $B \subseteq \mathcal{P}(V)$，满足 $|S| = 2$ 对所有 $S \in B$均成立）。从图形上看， G 表示为有向无环图（V, E），并额外包含双向边 $a \leftrightarrow b$，对所有 $\{a, b\} \in B$均成立。

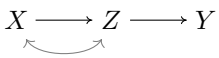


图 4：祖先有向混合图示例。

其思想是，双向边标示可能存在未包含在 V 中的潜在共同原因。假设图 5a 中的有向无环图 G 表示由因果充分集中的变量$V = O \cup L$ 构成的正确因果结构，其中 $O = \{X_1, X_2, X_3, X_4\}$ 且 $L = \{L_1, L_2\}$。有向无环图 G_O只是将 G 限制在 O 上，却遗漏了 L 对 O的因果相关性。一个顶点为 O、但能从G中“保留足够信息”（至少对于某些用途而言）的图，是将 G 潜在投影到 O 所得的祖先有向混合图（参见例如文献 [38]）。

定义 69.（潜在投影）[12]：给定一个 DAG $G = (V, E)$，其在 $O \subseteq V$ 上的潜在投影是祖先有向混合图 $\pi_O(G) = (O, E', B)$，其中

$$
\begin{aligned}
E' &:= \{(a, b) \in O \times O \mid \exists \textit{ directed path } a \to \ldots \to b \textit{ with all intermediate vertices in } V \setminus O\} \\
B &:= \{\{a, b\} \subseteq O \mid \exists \textit{ path of the form}^{13} \; a \leftarrow \ldots \to b \textit{ with all intermediate vertices in } V \setminus O\} \; .
\end{aligned}
$$

注意， E' 至少包含 E 中连接 O内顶点的所有有向边。图 5b 展示了图 5a 中 DAG 在 $O = \{X_1, X_2, X_3, X_4\}$上的潜在投影。

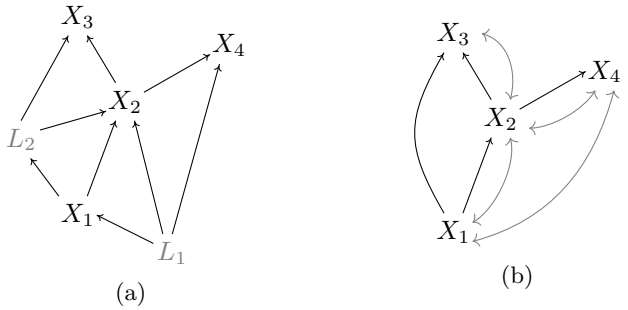


图 5 ：(a) 中的 DAG 示例及其在 (b) 中顶点子集上的潜在投影。

还有其他类型的图，能够将有向无环图的顶点投影到其子集，同时保留有向无环图的更多细节。特别是，边缘化有向无环图（mDAG）这一概念通过允许连接图中两个以上顶点的超边，推广了祖先有向混合图。Evans[91, 90], 引入这一概念，是为了研究由某个基于有向无环图的模型所蕴含的关于分布 P 的全部约束——包括不等式约束——

[11]在因果模型文献中，祖先有向混合图有时也称为“因果图”。[12]为简单起见，这里仅对 DAG 作出定义；将其扩展到祖先有向混合图在其顶点子集上的潜在投影则很直接 [38]。[13]路径，也就是一条 *not necessarily* 有向路径，是一系列边，不论其方向如何；这里的条件是两端的边分别指向 a 和 b。

由基于某个 DAG 的模型所隐含的 P。不过，本文不讨论 mDAG。

祖先有向混合图对于因果模型框架的重要性源于两个事实。首先，具有顶点 O 的祖先有向混合图A 编码了分布 $P(O)$ 上的所有等式约束，尤其是所有条件独立关系；若 $P(O)$ 是某个关于有向无环图 G 马尔可夫的分布的边缘分布，并且该图满足 $A = \pi_O(G)$，那么这些约束都必须成立。分布 $P(O)$ 相对于具有顶点 O 的祖先有向混合图 A 所满足的相容性条件，称为嵌套马尔可夫条件 [38]。其定义请参阅相关文献（参见文献 [38]），因为本文目前不需要用到该定义。其次，祖先有向混合图的重要性还在于，它们恰好编码了评估因果效应可识别性问题所需的充分细节；我们将在第7节探讨这些问题。

6.2 无向混合图与网络图

设 $\mathbb{M}$ 是一个因果贝叶斯网络，即根据定义2，是 FStoch 中的一个闭合因果模型，其因果结构为 $G_{\mathbb{M}} = (G, O)$。该模型定义了一个分布 $P(O)$，可由相应的网络图 $D_{\mathbb{M}}$ 表示，其输出为 O （将其解释为 FStoch 中的状态后）。根据构造，该分布 $P(O)$ 满足相对于祖先有向混合图 $\pi_O(G)$ 的嵌套马尔可夫条件；此图是将 G 潜在投影到 O 得到的。假设我们只有关于 O 的观察数据，但对潜在共同原因有具体了解，而这些信息编码在有向无环图 G 中。那么，网络图 $D_{\mathbb{M}}$ 确实是自然的研究对象。虽然图中的信道并非都能由关于 O 的观察数据唯一确定，但根据假设，这些机制确实存在，并且可以通过 $D_{\mathbb{M}}$ 优雅地进行推理。回想例33，其中（经解释的）网络图定义了一个状态（在 FStoch 中），其输出为 S、L 和 A，但机制 c_B 被明确纳入其中——尽管 B 可能无法观测到，该网络图仍表达了一个具体且合理的假设，即这四个变量之间存在相应的直接原因关系。

反过来，假设给定一个顶点为 O 的祖先有向混合图 G，以及一个分布 $P(O)$，它满足相对于 G 的嵌套马尔可夫条件。这正是因果模型文献中常见的因果查询问题所依据的给定数据。不过，既然状态网络图与基于有向无环图的因果结构一一对应（见命题 29），那么该如何将网络图与这些数据关联起来呢？事实上，如何以一种有原则的方式将给定的祖先有向混合图关联到一个有向无环图，有两种自然选择[14]。给定一个祖先有向混合图 $G = (V, E, B)$，我们考虑因果结构 (D, V)，其中 D 是以下有向无环图之一：

(1) $\rho(G) := (V \cup \{R_b\}_{b \in B}, E \cup E')$ 是通过为每条双向边 b 添加一个额外的根节点 R_b，并在满足 $b \in B$ 和 $v \in b\}$ 时添加从这些节点指向相应两个端点的边 $E' = \{R_b \to v \mid$，从而得到的有向无环图；

(2) $\tilde{\rho}(G) := (V \cup \{R_m\}_{m \in M}$、$E \cup E')$，其中 $M = \{m \subseteq V \mid m$ 是极大的，使得对于所有互不相同的 a、$b \in m$、$\{a, b\} \in B\}$ 和 $E' = \{R_m \to v \mid$，只要 $m \in M$ 且 $v \in m\}$，即通过为每个极大子集 $m \subseteq V$ 引入一个根节点 R_m 而得到，其中 m 中的所有顶点对在 G 中共享一条双向边，且 m 是其子节点。

图 6a 和图 6c 分别展示了图 6b 中祖先有向混合图 G 的诱导有向无环图 $\rho(G)$ 和 $\tilde{\rho}(G)$。

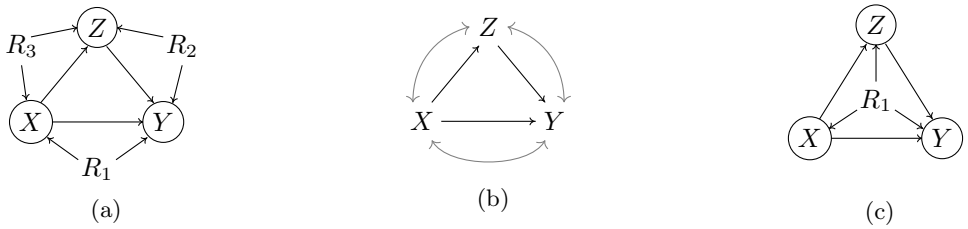


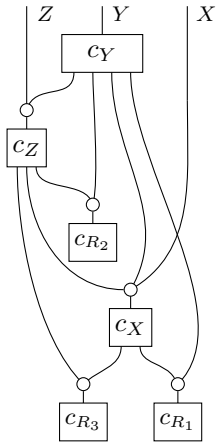
图 6：在 (b) 中的一个祖先有向混合图 G，以及其在 (a) 中的根化 $\rho(G)$ 和在 (c) 中的根化 $\tilde{\rho}(G)$ 示例。

[14] 对于第 6.1 节中提到的更一般的 mDAG，每个 mDAG [91 都有一个规范有向无环图].

我们将 $\rho\colon G \mapsto \rho(G)$ 和 $\tilde{\rho}\colon G \mapsto \tilde{\rho}(G)$ 称为根化映射——它们通过将根节点关联到双向边（的集合），把祖先有向混合图转换为有向无环图。

对于任意给定的祖先有向混合图 G，其诱导的因果结构 $\rho(G)$ 和 $\tilde{\rho}(G)$ 会诱导出相应的网络图 $D_{\rho(G)}$ 和 $D_{\tilde{\rho}(G)}$；将这些网络图在 FStoch 中解释后，便定义了相应的因果模型。还有许多其他有向无环图，它们的潜在投影都是 G，其诱导的网络图也同样可以纳入考虑。然而，重要的是，只要我们关注的是研究与给定祖先有向混合图 G 相关的 $P(O)$ 上的等式约束，或是涉及 O 中变量的因果效应的可识别性问题（见第7节），那么选择哪一个这样的有向无环图及其相应的网络图并不重要，前提是该有向无环图在 O 上的潜在投影确实是 G [38]。

因此，许多因果推断问题都可以通过图示来研究：从某个祖先有向混合图 G 转到由根化映射 ρ（或 $\tilde{\rho}$）诱导的网络图。后者是一种规则，能够根据任意给定的祖先有向混合图唯一确定一个有向无环图。随后，在 FStoch 中解释 $D_{\rho(G)}$，意味着将根化引入的附加根节点视为与 O 同类的、未具体指定且不受约束的变量，即有限离散变量。同样，$D_{\rho(G)}$ 中出现的信道只需视为任意信道，只要对附加根节点进行边缘化后能还原分布 $P(O)$ 即可；而信道的具体形式通常无关紧要。例如，我们可以通过图 6a 中的根化 $\rho(G)$，将式（44）所示形式的因果模型与图 6b 中的祖先有向混合图 G 联系起来。



(44)

总之，给定一个具有有向无环图 G 的因果贝叶斯网络 $\mathbb{M}$，我们可以针对某个顶点子集 O，考察其诱导出的祖先有向混合图 $\pi_O(G)$；但表示实际因果结构的是有向无环图，也就是网络图 $D_{\mathbb{M}}$。相比之下，给定一个顶点为 O 的祖先有向混合图 G'，以及某个与之相容的 $P(O)$，我们可以考察通过 $\rho(G')$ 诱导出的有向无环图及相应的因果模型，以便针对给定数据进行图示推理；但这并不意味着该有向无环图就是因果结构。相反，它是一种具有恰当特征、可用于某些目的下进行推理的合理因果结构 $D_{\rho(G')}$。给定数据的背景总能明确一个因果模型应被赋予何种地位。

7 因果效应的识别

7.1 问题

在许多情况下，人们都想知道干预的效应。然而，往往无法通过实验获取这一信息，也就是实际实施干预——这可能在技术上不可行（例如固定行星的位置或人们的社会经济地位），也可能在伦理上无法接受（例如强迫人们吸烟）。Pearl 将 $P(Y;\ \mathrm{do}(X))$ [15] 称为 X 对 Y 的因果效应。因此，一个重要的问题是：根据何种数据、在什么情况下，我们才能预测如果对某些变量进行干预会发生什么，例如确定因果效应 $P(Y;\ \mathrm{do}(X))$？

仅凭观察数据，也就是某个分布 $P(O)$，无法对任何干预后分布作出任何判断。另一方面，如果已知变量 V 上的因果模型 $\mathbb{M}$，其中包含每个变量 P 的因果机制，

[15]回想一下，这是 $\mathrm{do}(X = x)$ 中不同取值 x 所对应的干预后分布族；等价地，它是由相应的开放因果模型所定义的信道，正如第5节所述。

并且已知任意干预（定义 43），无论是否为原子干预，那么描述干预后情景的结果模型也就完全确定了。具体而言，任意变量子集上的边缘分布都是已知的。对于原子干预 do(X)，这称为“截断因子分解公式” [4]（参见第 4 节的讨论）。

若某个量 ϕ 相对于某些数据 E 可识别，当且仅当所有在 E上达成一致的因果模型也必然在 ϕ上达成一致。如果给定数据由 $P(O)$构成，且只有部分因果知识，那么理解因果效应何时以及如何可识别，就是因果效应可识别性问题。更精确地说，该问题定义如下：给定 $P(O)$、一个顶点为 O、包含不相交子集 X、Y、$Z \subset O$的祖先有向混合图 G，何时$P(Y|Z; \text{do}(X))$可识别？[16]

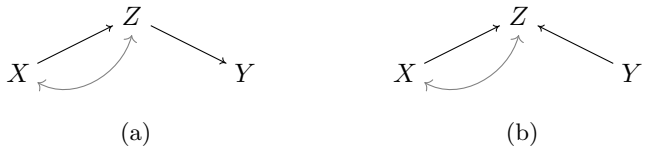


图 7：来自 [41] 的因果效应可识别性示例。如果给定 $P(X, Y, Z)$，那么对于 (a) 中的祖先有向混合图，$P(Y|Z; \text{do}(X))$ 可识别，而 $P(Y; \text{do}(X))$ 不可识别；对于 (b) 中的祖先有向混合图，情况则相反。请注意，由于未满足第 7.2节中的 c-分量判据，$P(Y, Z; \text{do}(X))$ 在两种情况下都不可识别。关于这两种情况的图解处理，请参见第 7.3节。

因果效应可识别性问题在过去 30 年中得到了深入研究，并且此后已经得到解决。例如，参见 Shpitser 和 Pearl [41] 的引言和综述，其中包括对该问题的可靠且完备的算法解法的介绍。有关几个教学示例，请参见图 7。

更一般地说，如果给定 $P(O)$ 和一个顶点为 O 的祖先有向混合图 G（也可能还给定其他一些干预后分布），人们可以询问某种一般的非原子干预 σ_X 下 $P(Y|Z; \sigma_X)$ 的可识别性。Correa 和 Bareinboim 在文献 [37], 中定义并研究了这一更一般的问题，并提出了一种算法解法，证明了该解法是可靠的。

现在，既然这个问题已经得到解决——存在算法解法，并且关于元组 (G, X, Y, Z) 可识别性的必要且充分的图形条件也已知——为什么还要在这里花费篇幅讨论它呢？首先，这些必要且充分的图形条件并不特别容易判断（需要考察是否不存在所谓的 hedge [41], ，或者等价地，考察某些区域 [38]是否具有“内在性”）。其次，即使已经知道某个问题实例是可识别的，算法解法仍需应用一系列 do-演算规则 [41]（或者更一般地，对于非原子干预 [38], 应用 σ-演算规则），而知道哪些规则能够将我们引向解答并不总是容易。因此，尽管算法带来了自动化，如果存在一种以图示形式表示该问题、从而更直观且更易于推理的方法，那么无论对于希望用纸笔分析具体案例的实践者，还是出于教学目的，这种方法都会很有价值。在接下来的两个小节中，我们呼应 Jacobs 等人在文献 [9], 中的工作，论证这确实是可行的。

7.2 c-分量条件的图示表达——JKZ 结果

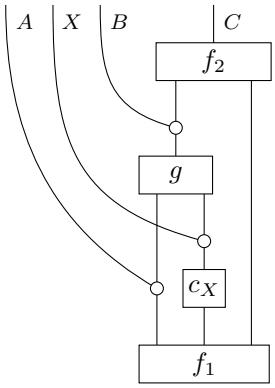
对于 $|X| = 1$ 和 $Y = V \setminus X$，会出现这一问题的一个特殊情形，即干预只涉及一个变量，而我们要考察 $P(V \setminus X; \text{do}(X))$ 的可识别性；这是所有其他变量上的干预后分布。Tian 和 Pearl 指出，在这种情况下，存在一个简单的图形条件，且该条件是可识别性的充要条件 [79]。给定一个祖先有向混合图 $G = (V, E, B)$，对于顶点 X，其 c-分量为 $C(X)$: $= \{v \in V \mid \exists$ 中 X 与 $v\}$之间的混杂路径，其中混杂路径是仅由双向边构成的路径。Tian 和 Pearl 提出的条件，即 c-分量条件，可表述为$Ch(X) \cap C(X) = \emptyset$，也就是说，$X$ 与其任何子节点之间都不能存在混杂路径。

Jacobs、Kissinger 和 Zanasi 在参考文献 [9] 中证明，c-分量条件有一个简洁的等价 t 表述形式，是一种结构 网络图上的结构约束，而且解决方案，即表达 g

[16]集合 Z 当然可能为空，但请注意，一般而言，可识别性取决于查询是否涉及某些作为条件的变量—— $P(Y|Z; \text{do}(X))$ 的可识别性并不等价于 $P(Y, Z; \text{do}(X))$ 或 $P(Y; \text{do}(X))$ 的可识别性。此外，尽管文献中通常对此不作明示，但当然假设 $P(O)$ 与给定的祖先有向混合图 G相容，也就是说，它满足嵌套马尔可夫条件。

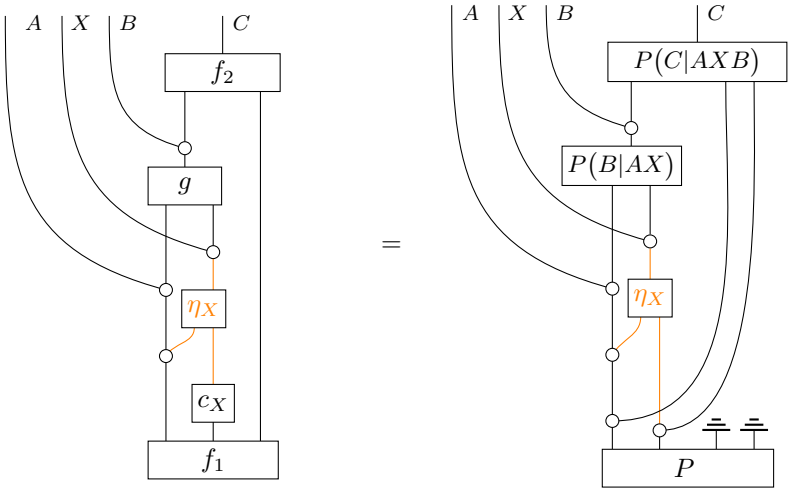
在给定分布 $P(O)$ 的条件下，$P(V \setminus X; \mathrm{do}(X))$ 也可以通过图示地推导得到。我们在此重述这一结果，以供后续参考，并将其稍作推广，因为当前的框架允许以图示方式处理任意干预，而不再局限于 [9] 中对切断干预的处理。

定理 70。（JKZ 定理 [9][17]）：设 D 是 FStoch 中定义在变量 V 上、输出为 $O \subseteq V$的网络图。设 $X \in O$，并假设存在一种将其余变量划分为不相交子集 $O \setminus \{X\} = A \cup B \cup C$ 的方式，使得 D 可以写成如下形式（也就是说，f_1、f_2、g 是网络图）：[18]



(45)

那么，对于如下形式的干预 η_X，$P(O; \eta_X)$ 可由任何能够从 FStoch 中对 D 的解释得到的 $P(O)$ 识别，并且其表达式为：



(46)

至关重要的是，式（46）的右侧只包含边缘分布和条件分布；忽略全支撑问题时，它们均可由 $P(O)$ 计算得到。注意，$\mathrm{do}(X = x)$ 干预这一特殊情况可通过令 $\eta_X = x \circ (\doteq_A \otimes \doteq_X)$ 得到。[19]

正如 Jacobs 等人在文献 [9] 中所论证的（用我们的术语来说），给定一个顶点为 O 的祖先有向混合图 G，若对某个$X \in O$，有 $Ch(X) \cap C(X) = \emptyset$，那么 $O \setminus X$ 可以划分为互不相交的子集 A、B、C，使得由 G 诱导的网络图 $D_{\rho(G)}$ 具有形式 (45)。反过来，从一个具有这种形式的网络图D 出发，不难看出，$\pi_O(G_D)$——由 D 诱导的有向无环图到 O 的潜在投影 —— 满足关于 X 的相应 c- 分量条件。这种等价关系背后的直觉是，任何连接到 X 的混杂路径——也就是说，未被复制到输出端的导线——都可能连接到 X 的后代，但不会连接到 X 的子节点；在式 (45) 中，这些子节点必须包含在 B 中。

[17]这一结果的证明只依赖于“梳状析解”的唯一性，该唯一性也已在参考文献 [9] 中证明；因此，对于更一般干预的当前陈述，其证明与参考文献 [9]中证明的版本完全相同。[18]等价地，f_1、f_2、g 是 Free(D) 中的态射，参见注释 35。[19]当然还有修剪，参见注释 46。

注记 71。请注意，与文献 *[9],* 中的说法相反，定理 70 并不比 Tian 和 Pearl 的原始结果更一般，因为前者并未涵盖超出后者所依据的“半马尔可夫性”假设的情形。半马尔可夫性是关于未观测变量性质的一项假设，即这些变量只能是根节点。[20] 这是一个微妙的问题，涉及潜在投影的作用，以及半马尔可夫性作为因果效应可识别性的一项假设，其地位在文献中似乎并非始终清晰。在文献 *[38]* 中，Richardson 等人证明了一项有助于澄清这一问题的结果——详情请见下文的定理 72 及后续章节中的讨论。

7.3 一般情形的图示——示例

除了适用 c-分量条件的情形之外，让我们回到因果效应 $P(Y|Z;\mathrm{do}(X))$ 可识别性的一般情形：对于任意不相交子集 $X, Y, Z \subset O$，给定 $P(O)$ 和一个顶点为 O 的祖先有向混合图 G。在这个一般情形下，我们也已知可识别性的一个充要图形条件，因此不妨思考一下是否存在与之对应的定理 70 泛化。然而，该图形条件——不存在所谓的 hedge [41][21] ——是一个相当复杂的图论概念，我们在此不作讨论，而是请读者参阅相关文献。不过我们注意到，仔细考察其定义后，显然不存在一种统一形式，能让网络图在该条件成立时都化为这种形式；这与定理 70 中的 c-分量条件不同。[22]

重要的是，即使目前还没有能够推广定理 70 的一般性结果，图示方法对于研究因果效应的可识别性仍然大有裨益。在这一情境中，网络图是一种尤其自然且合适的推理语言；它们将因果知识与概率论中的规则结合起来，使许多原本看似复杂、难以察觉的步骤变得相当直接。为印证这一看法，下面将讨论四个例子，它们此前都曾以传统的、也就是非图示的方式在文献中得到处理：例 73 和例 74 是图 7 中的简单例子，也作为教学示例出现在参考文献 [41] 中；例 75 是参考文献 [41] 中一个较复杂的例子；例 76 出自参考文献 [37]，涉及非原子干预。这些例子均不在定理 70 的涵盖范围内。

结合第 6 节的内容，以图示方式讨论这些例子时，将遵循预期的方法。给定一个顶点为 O、包含三个互不相交子集 X、Y、$Z \subseteq O$ 的祖先有向混合图 G，考虑其根化 $\rho(G)$ 以及相应的诱导网络图 $D_{\rho(G)}$。接着，令 $\mathbb{M}$ 为一个相应的因果贝叶斯网络，其具有 $D_{\rho(G)}$，并在 C 中有一个解释 $=\mathrm{Mat}_{\mathbb{R}^+}$；再考虑基于如此定义的因果模型的信道 $P(Y|Z;\mathrm{do}(X))$，其输出态为 $P(O)$。目标是简化并改写 $P(Y|Z;\mathrm{do}(X))$，使其能够用 $P(O)$ 表示，也就是用 $P(O)$ 本身及其边缘分布和条件分布来表示。接下来可能出现三种情况，揭示出句法与语义之间的重要相互作用：

(1) 成功将 $P(Y|Z;\mathrm{do}(X))$ 改写为关于 $P(O)$ 的表达式，而且只用到了任何具有效应条件化的 cd-范畴中都成立的图示推理，因此既不依赖具体的范畴 C，也不依赖 $P(O)$ 的偶然属性。由此可得可识别性。

(2) 成功将 $P(Y|Z;\mathrm{do}(X))$ 改写为关于 $P(O)$ 的表达式，但前提是对特定 $P(O)$ 的属性作出一些假设（例如某些标量是否为零，见下文示例）。同样，尽管现在需要以 $P(O)$ 的偶然属性为前提，仍可得出可识别性。

(3) 无法将 $P(Y|Z;\mathrm{do}(X))$ 改写为关于 $P(O)$ 的表达式。因此，基于简化图论证不可识别性，所依赖的假设比 C 是一个带有效应的 cd-范畴更强

[20]有时，半马尔可夫性被理解为这样一个条件：未观测到的根节点至多还有两个子节点。根据定理 72，只要给定的祖先有向混合图被视为因果充分变量集的正确因果结构的潜在投影，那么无论较弱还是较强的半马尔可夫性概念，都不是因果效应可识别性结果所需的假设。[21]等价地说，[38] 可用某些“区域”的内在性来表述这一图形条件。[22]无论如何，人们或许会想探究它与 d-*separation* 的图形条件之间是否存在一种可能颇有启发性的类比（见第 3.4 节和附录C.5）。后者同样不存在一种简单、单一的等价图示条件，能将其表述为由有向无环图诱导的网络图上的某种特定结构约束。不过，Fritz 最近 *et al.* 证明，确实存在一个与 d-分离等价的简洁条件，但它是通过对网络图进行手术操作来表述的，而非针对图本身的某一种特定形式 [12]。类似地，不难设想，若允许对网络图进行诸如“手术”之类的图示操作，就有可能找到一个在一般情形下与因果效应可识别性等价的图示条件。我们将这一可能性留待未来工作探索。

条件化——本质上，需要范畴中有足够多的映射，才能确保构造出合适的反例。下文有两个被认为不可识别的例子——我们在 $\mathsf{C} = \mathsf{Mat}_{\mathbb{R}^+}$ 中理解这一结论；除了给出直观理由外，我们仅引用文献中的合适反例。

现在，我们之所以能在重写成功时得出可识别性结论，即情形 (1) 和 (2)，要归功于 Richardson 等人在参考文献 [38]（其中的推论 50）中提出的以下结果——他们称之为“民间结果”，因为尽管它十分重要，且在许多早期研究中被隐含地使用过，但此前从未有人正式证明：

定理 72. ([38]): 设 G 和 G' 是两个有向无环图，其各自的顶点集合 V_G 和 $V_{G'}$ 在某个集合 O 上重叠，即 $O \subseteq V_G \cap V_{G'}$，并且它们到 O 的潜在投影相同，$\pi_O(G) = \pi_O(G')$。那么，对于不相交子集 $X, Y \subset O$，若假设 G 是因果结构，则因果效应 $P(Y; \mathrm{do}(X))$ 可从 $P(O)$ 中识别，当且仅当假设 G' 是因果结构时，它可从 $P(O)$ 中识别；此外，若该效应确实可识别，则两个有向无环图都蕴含相同的 $P(Y; \mathrm{do}(X))$。

这一结果对本文的意义当然在于：将祖先有向混合图通过根化（见第 6.2 节）转化为网络图，再基于该网络图进行推理，这一过程不会损害相应因果效应的可识别性，也不会限制一般性。[23]

正如概率记法所示，我们考虑的是 $\mathsf{C} = \mathsf{Mat}_{\mathbb{R}^+}$，但也可以分别将 $P(O)$ 和 $P(Y|Z; \mathrm{do}(X))$ 视为开放因果模型的输出上的状态和信道；它们由某个具有效应条件化的 cd-范畴 C 中相应的因果模型所定义。只要 C 满足定理 72 能够适当地推广，使得考虑根化 $\rho(G)$ 不会引入除给定祖先有向混合图 G 之外的任何因果假设，那么我们的方法就能让人们在任意这样的 C 中，针对情形 (1) 和 (2) 得出可识别性的结论。

例 73。考虑图 8a 中的祖先有向混合图 G（为方便起见，此处从图 7a 重绘；另请参阅参考文献 *[41]*，了解对该例子的讨论），以及图 8b 中对它的根化 ρ（G）和图 8c 中对应的网络图 $D_{\rho(G)}$。

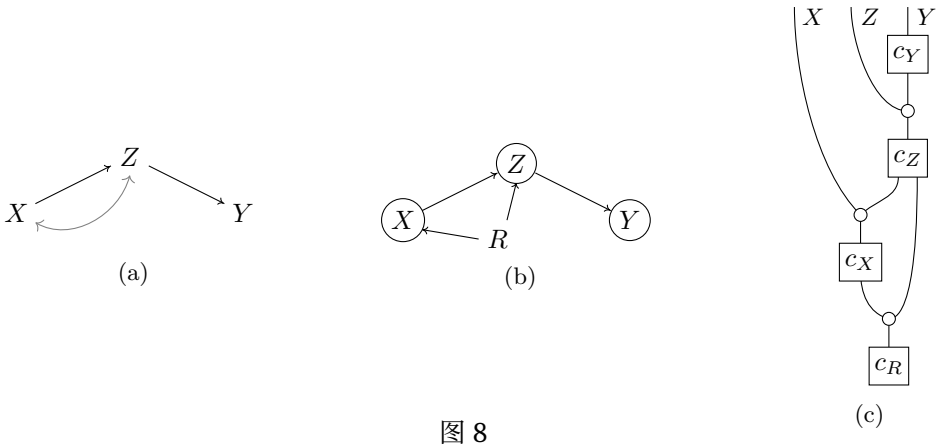
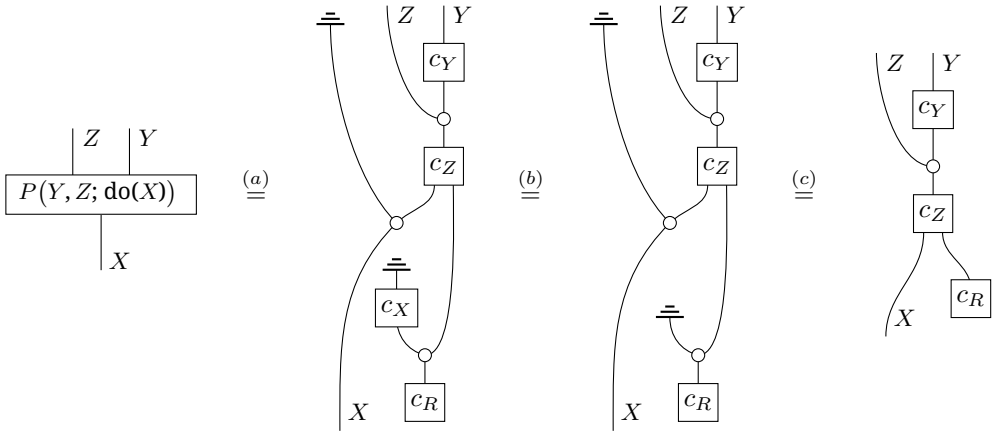


图 8

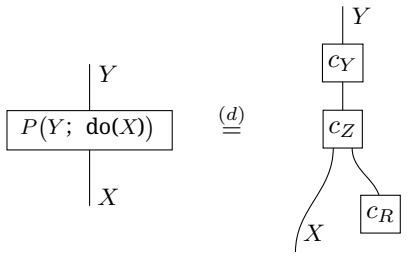
根据 $D_{\rho(G)}$ 进行推理，让我们研究图 7 图注中的论断：$P(Y \mid Z; \mathrm{do}(X))$

[23]请注意，对于由不相交子集 X、Y、$Z \subset O$构成的 *triple*，形如 $P(Y|Z = z; \mathrm{do}(X))$ 的因果效应的可识别性，可归结为针对一组新顶点集合 *pair* 的“无条件”情形；这组集合根据集合 X、Y、Z 定义，且重要的是，相对于同一个祖先有向混合图 G [41]。因此，尽管定理 72 的结果只提及“无条件”因果效应，但仍可据此得出结论：通过对给定的祖先有向混合图进行根化来研究 $P(Y|Z = z; \mathrm{do}(X))$ 的可识别性，并不失一般性。

是可识别的，而 $P(Y;\,\mathrm{do}(X))$ 则不可识别。首先，注意：

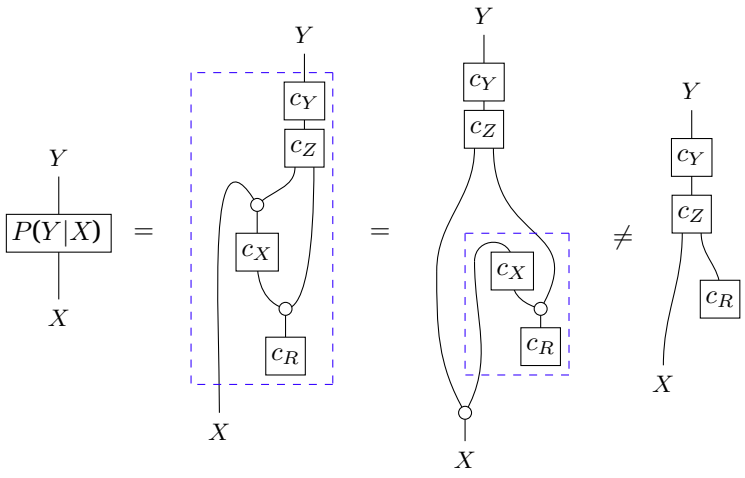


步骤 (a) 只是应用 $\mathrm{do}(X)$ 的定义，即定义一个相应的开放模型（不输入特定状态 x），参见第5节；步骤 (b) 仅使用了 c_X 是一个信道这一事实；步骤 (c) 则利用了复制与丢弃之间的相互作用（参见定义2）。接下来，只需对 Z 进行边缘化，便可得到：

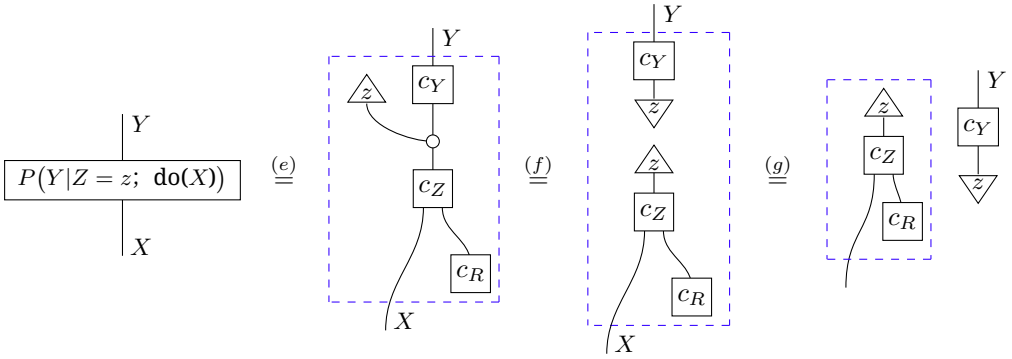


显然，由于存在 R，无法从 P（X，Y，Z）中识别出这一点。我们不再重复标准证明，因为这是 X 与 Z 之间存在未观测混杂因素 R 的典型例子，而是说明如何从图示中直观地看出这一点。一般而言，对 c_Z 和 c_R 的不同选择显然会影响 $P(Y;\,\mathrm{do}(X))$。然而，正如图 8c 的网络图所示，可以对这两个信道进行精细调节，使其不影响 P（X，Y，Z），因为这种差异会由对信道 c_X 的适当选择予以补偿。

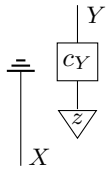
还需注意，由提取出的（部分）信道 $P(Y|X)$ 给出的贝叶斯条件分布（见第2.5节）是可识别的，但通常与所求信道不同。下面的简化采用了贯穿第2.5节和附录C的标准步骤：



接下来，考虑条件因果效应

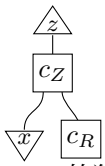


步骤 (e) 使用了条件分布的图示定义（第2.5节）；步骤 (f) 使用了式 (6)；步骤 (g) 使用了式 (11)，并利用了 c_Y 和 z 已经归一化这一点。现在请注意，如果归一化框中的效应具有全支撑，那么归一化框就会将该效应直接变为丢弃，因此，所关注的信道 $P\big(Y|Z=z;\,\mathrm{do}(X)\big)$ 等于



在这种情况下，条件因果效应 $P\big(Y|Z=z;\ \mathrm{do}(X=x)\big)$ 独立于 x，并且由于信道 c_Y与贝叶斯条件分布 $P(Y|Z)$ 相同（在其有定义的地方），因果效应 $P\big(Y|Z=z;\ \mathrm{do}(X)\big)$ 是可识别的，也就是说，只要 $P(Z=z)\neq 0$，它就由 $P\big(Y|Z=z\big)$ 给出。文献 *[41]* 中也得到了这一答案。观察图 8a 中的祖先有向混合图 G，这显然是意料之中的，因为 Y 仅通过 Z依赖于 X，而这里对 Z进行了条件化，条件为 $Z=z$。

然而，在关注条件分布时，两种情况之间存在一个关键区别。一种情况是断言某个特定状态可识别，但根据给定数据的具体情况，它可能是零态；在文献中，这通常会被称为条件分布未定义（参见第2.5节）。另一种情况则是它不可识别。在上面的例子中，$P\big(Y|Z=z;\ \mathrm{do}(X=x)\big)$似乎是可识别的，但严格来说并非如此。标量是否



对所有 x– 都不为零，只有这样才能发现 $P\big(Y|Z=z$ 的独立性；$\mathrm{do}(X=x)\big)$ 来自 x– 取决于一个实际上不可识别的量，也就是说，正是出于导致 $P\big(Y$ 的同一原因；$\mathrm{do}(X)\big)$ 不可识别。这是一个微妙之处，很容易被忽略，正如 Ref. *[41]* 中似乎就是如此——图解处理有助于防止这种情况。

例74。考虑图9a 中的祖先有向混合图 G（为方便起见，此处再次从图7b复制；有关此例的讨论，参见文献 *[41]*），以及图9b 中的根化 ρ（G）和图9c 中对应的网络图 $D_{\rho(G)}$。

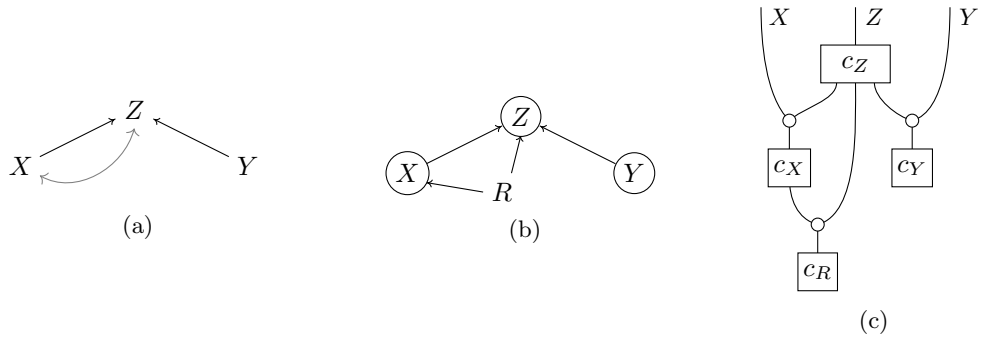
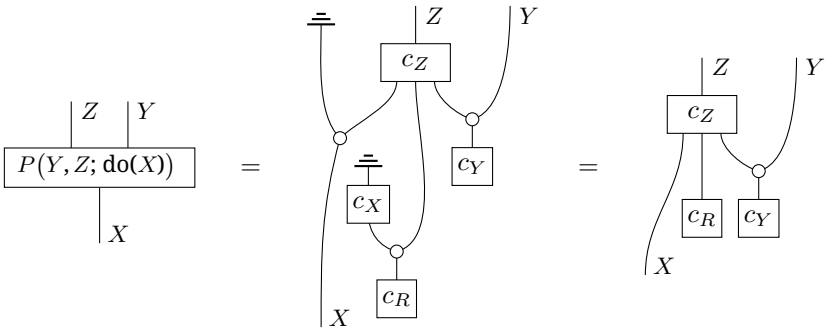
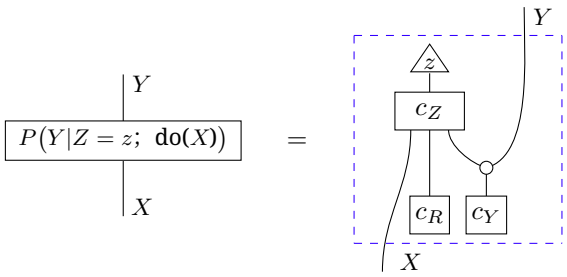


图9

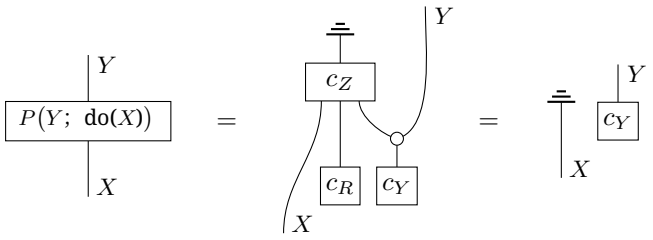
通过对 $D_{\rho(G)}$ 进行推理，可以再次轻松地研究图 7 图注中的论断：P（$Y \mid Z$；do(X)) 不可识别，而 P（Y；do(X)) 可识别。以下步骤与例 73 类似。



因此，对 $Z = z$进行条件化后，得到的表达式包含无法从 $P(X, Y, Z)$中识别的量：[24]



相比之下，对 Z进行边缘化会使 Y与对 X和 $P\big(Y$的任何依赖分离；do(X)$\big)$ 最终是可识别的——它等于边缘分布 $P(Y)$：



请注意，与上面的例 73 不同，最后一步中的这个等式无条件成立，而且在这种情况下，可识别性并无微妙之处。

[24]我们再次不证明这一事实；类似的直观论证见例 73。

例 75. 考虑图 10a 中的祖先有向混合图，并假设给定了 $P(Y_1, Y_2, W_1, W_2, X)$。那么 $P(Y_1, Y_2; \text{do}(X))$ 是否可识别? Shpitser 和 Pearl 在参考文献 *[41]* 中讨论了这个例子，并使用 IDENTIFY算法肯定地回答了这个问题。

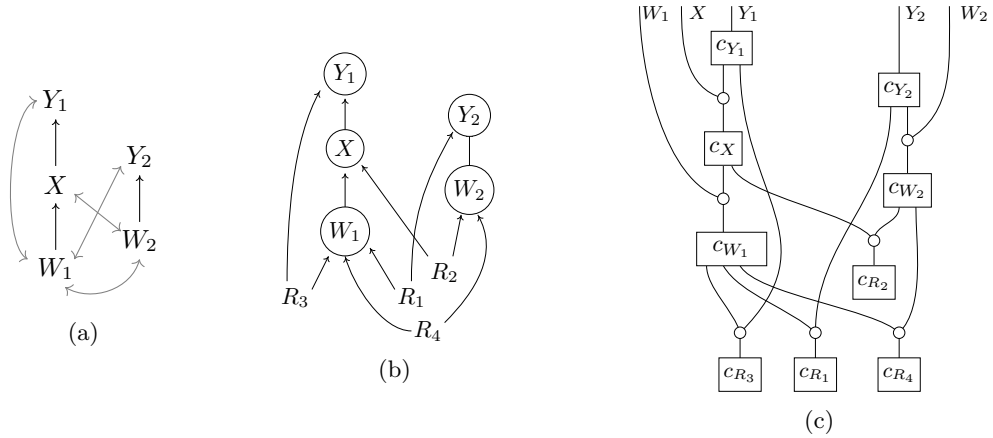
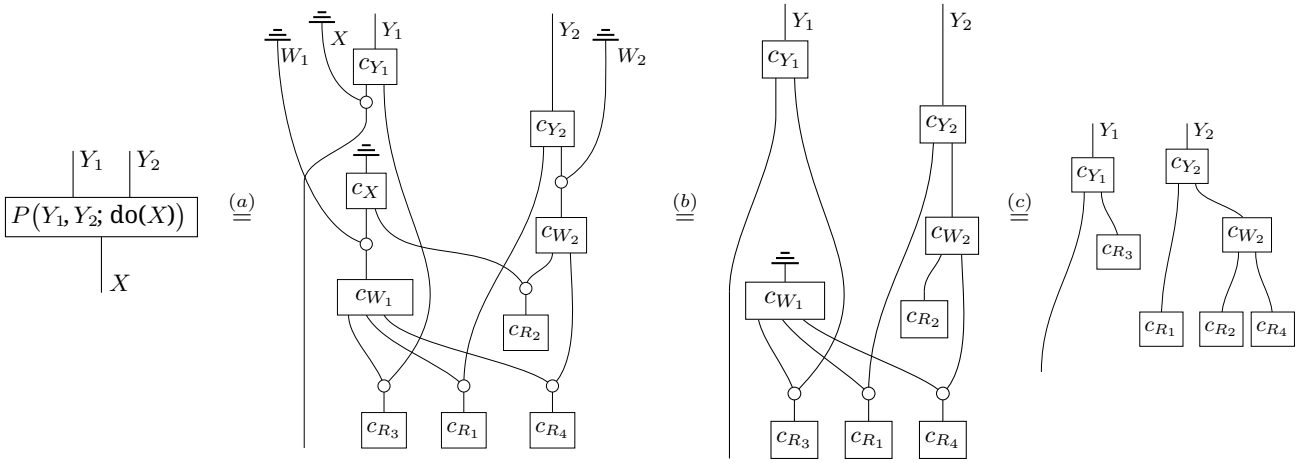


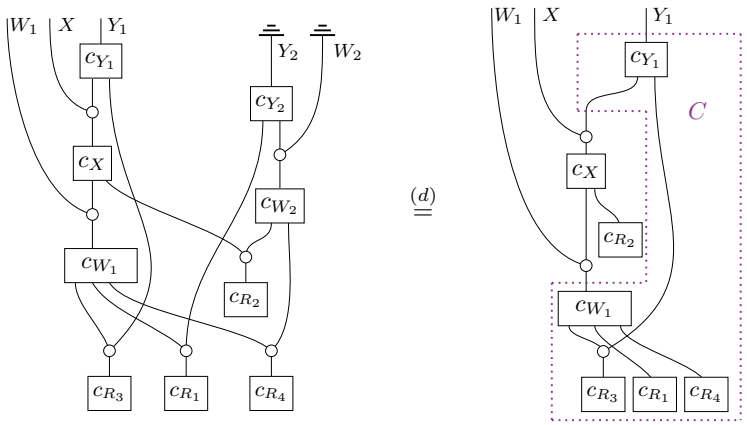
图10

考虑由图 10b 中的根化 $\rho(G)$ 所诱导的图 10c 中的网络图 $D_{\rho(G)}$。直接计算可得:

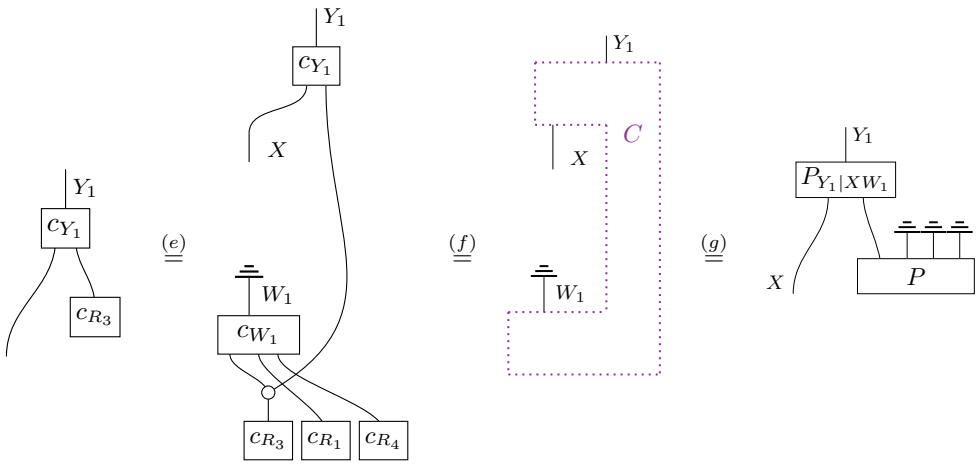


其中，步骤（a）只是对 $D_{\rho(G)}$ 应用相应的开放操作（和边缘化），而（b）和（c）是显而易见的简化（参见例 73 中的类似步骤）。因此， $P(Y_1, Y_2; \text{do}(X))$ 的可识别性问题归结为（c）右侧两个因子的可识别性。第二个因子正是边缘分布 $P(Y_2)$，这一点很容易通过丢弃 $D_{\rho(G)}$ 中的所有输出（除.....之外）看出

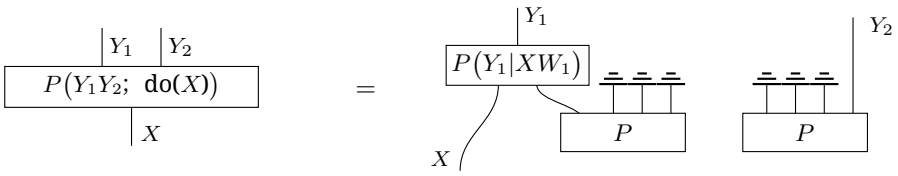
Y_2。但第一个因子呢？为了从 P 中“提取”它，考虑边缘分布 $P(Y_1, W_1, X)$：



其中 (d) 只进行了一如既往的简化操作，包括丢弃操作和移动方框。现在请注意，只要该分布具有全支撑，就可以通过梳状分解 *[9]*，从 $P(Y_1, W_1, X)$ 中识别出紫色虚线方框，而这又得出



其中 (e) 取了 (d) 的左侧，并展开了 X，同时对除 Y_1 之外的所有输出取边缘分布；步骤 (f) 仅使用了 (d)，而步骤 (g) 使用了参考文献 *[9]* 中的上述结果。因此，我们可以得出可识别性，其识别表达式为：



这与参考文献 *[41]* 中的答案一致。

作为最后一个例子，我们假设 C 中的帽是可消去的（2.3），$\mathsf{Mat}_{\mathbb{R}^+}$ 中的情况正是如此。

例 76。作为一个涉及非原子干预的例子，我们来看下面这个例子，它取自参考文献 *[37]*。设 G 是图 11a 中的祖先有向混合图，并给定 $P(R, W, X, Y, Z)$，以及 $P(R, W, X, Y, Z;\ \sigma_Z)$，其中干预 σ_Z 将 c_Z 替换为某个 $c_Z^*: X \to Z$。现在，令 σ_X 为将 c_X 替换为某个机制 $c_X^*: R \to X$ 的干预，那么 $P(Y|R, Z;\ \sigma_X)$ 能否根据给定数据识别？注意，在此情形下，给定数据不仅包括观察数据，还包括关于 σ_Z 的干预数据。

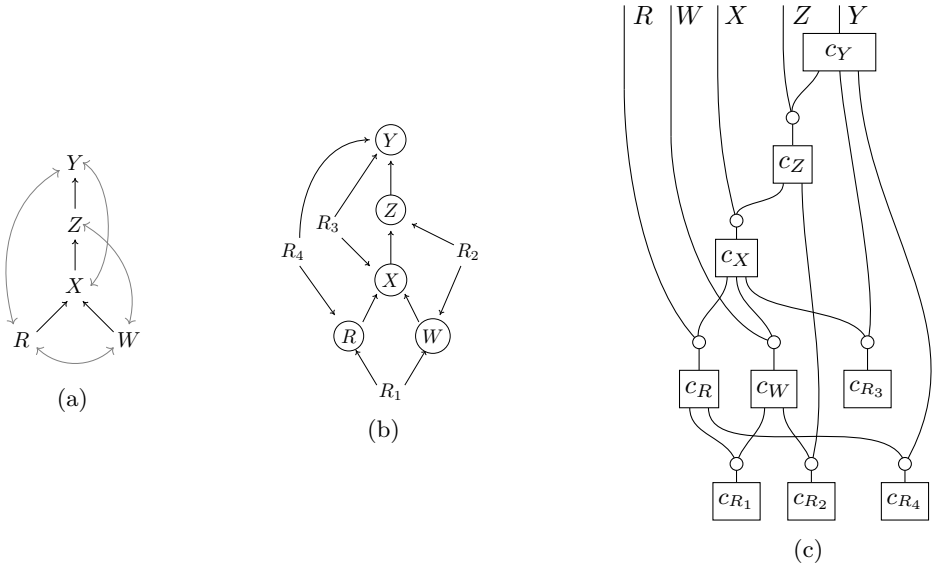
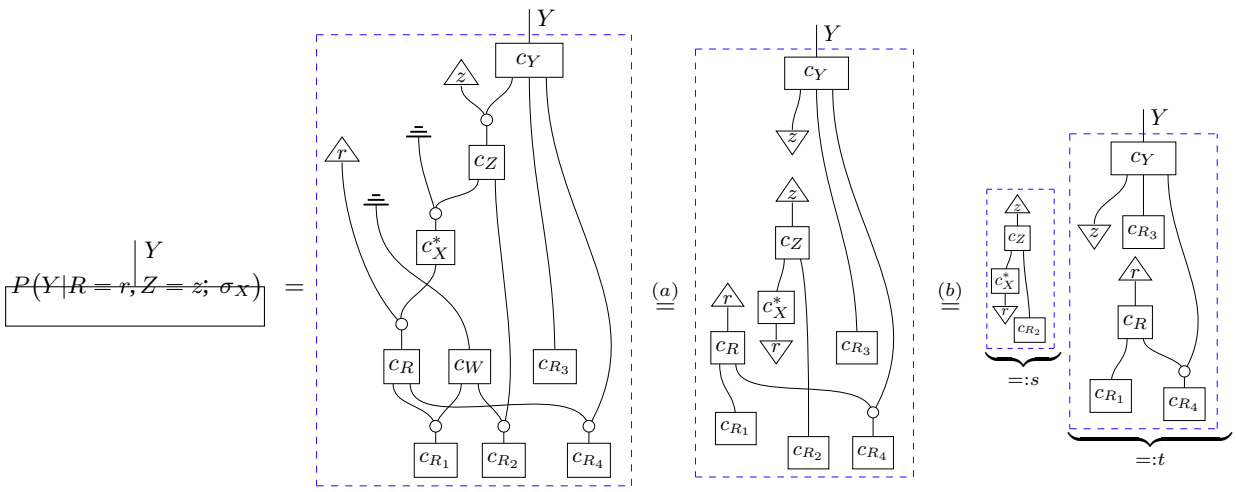


图 11

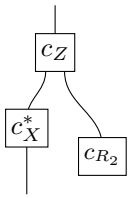
根据图 11b 中的根化 $\rho(G)$ 所诱导的图 11c 中的网络图 $D_{\rho(G)}$，[25] 感兴趣的量可以表示为：



其中 (a) 和 (b) 只需进行通常的简化。

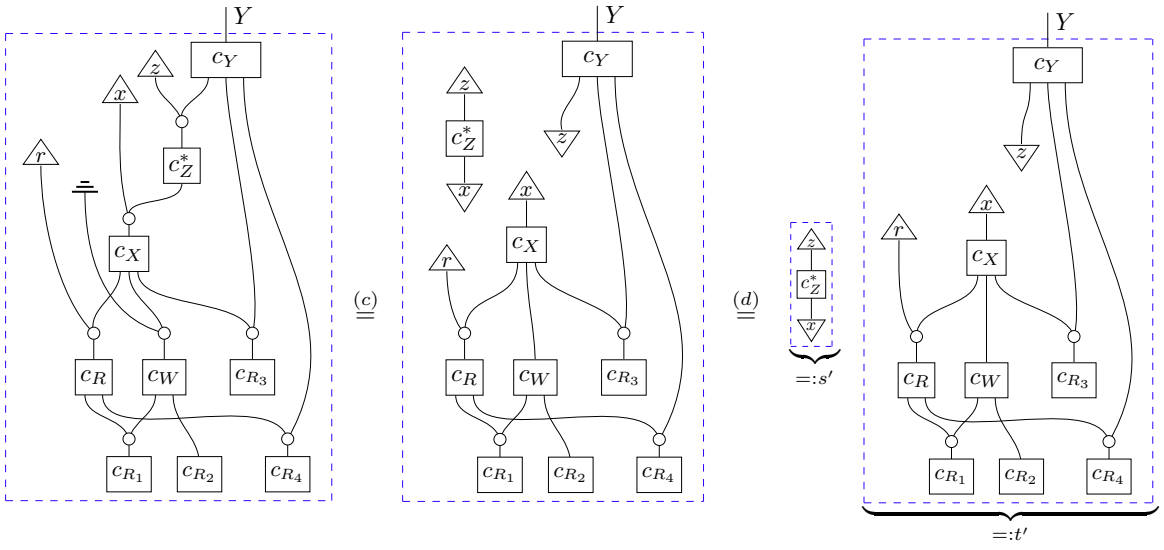
首先请注意，与例 73 类似，这里存在一个与标量可识别性有关的微妙之处。只要 s 的归一化框内的标量非零，归一化后的 s 就是实数 1，因此可以省略。然而，如果该标量为零，归一化后的 s 仍然为零，显然不能省略。要判断属于哪种情况，需要知道其身份

[25]请注意，定理 72 并不适用于比 do-干预更一般的干预，至少不能直接适用。据我们所知，尚未证明存在类似的「民间结果」，即在一般干预下，潜在投影 ADMG 对因果效应可识别性而言也不失一般性。不过，对于这个特定例子，不失一般性。这是因为，任何以 G 为其潜在投影的有向无环图，都是 $\rho(G)$（$= \tilde{\rho}(G)$）的一个特例，并且确实是一个可依据 $D_{\rho(G)}$ 识别的例子。



虽然根据假设， c_X^* 已知，但 c_Z 和 c_{R_2} 的复合实际上无法从给定数据中识别（参见例 73 中类似的论证）。因此，严格来说， $P(Y|R,Z;\sigma_X)$ 也不可识别——这又是一个在图解处理中很容易发现的微妙之处（而参考文献 *[37]* 中似乎忽略了这一点）。假设 c_X^*、 c_Z 和 c_{R_2} ——无论它们的确切身份为何——满足对于所有 z 和 r， s 都等于 1，那么就可以继续评估剩余状态 t 的可识别性。

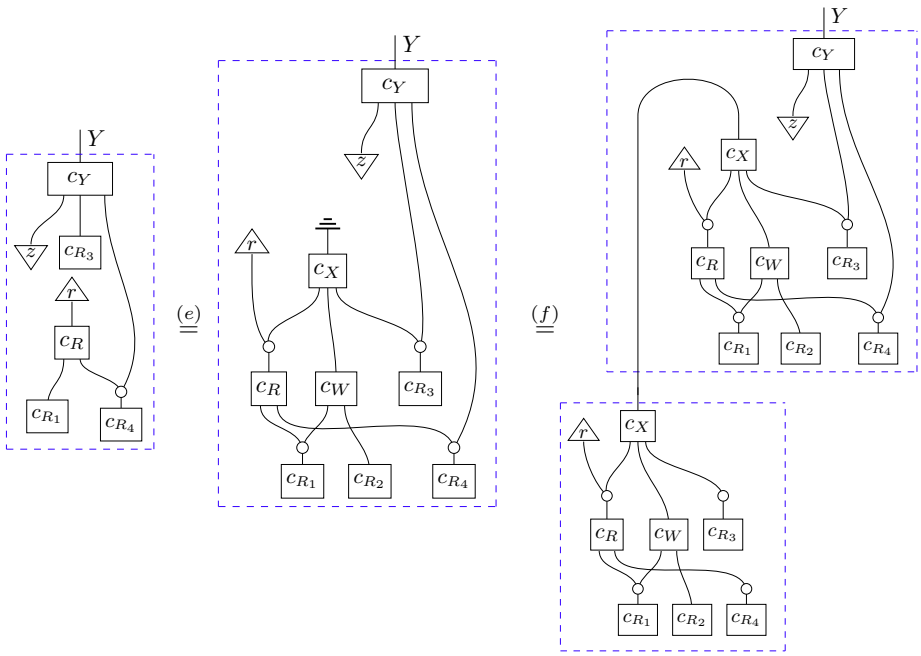
现在，为了判断给定的数据能否识别出该状态 t ，考虑给定的分布$P(R,W,X,Y,Z;\sigma_Z)$。由于 σ_Z的 c_Z^*并未出现在 t中，目标就是消除它的影响。因此，我们来考察条件分布 $P(Y|R=r,X=x,Z=z;\sigma_Z)$，以便如下所示将信道 c_Z^*“分离”出来:



其中，(*c*) 和 (*d*) 使用了方程（6）和（11）。同样，只要满足 $z\circ c_Z^*\circ x\neq 0$，就可以省略标量$c_Z^*$。根据假设，$c_Z^*$ 已知，因此总是可以计算 $z\circ c_Z^*\circ x$。

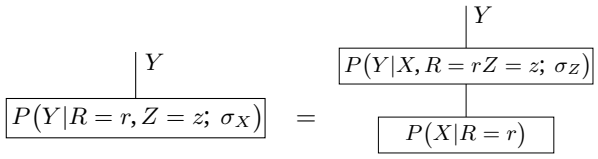
为了了解状态 t 与 t' 的状态之间的关系，请注意，当然可以“复制后丢弃”。后者是指用复制映射替换一条导线，并再次丢弃其一个输出，如前面所做的那样（多次

在下文步骤 (e) 中乘以的次数。



步骤 (*f*) 只是应用图示链式法则（引理 12）。

总之，若不作出关于 s 的假设，也就是对 c^*_X、c_Z 和 c_{R_2} 作出相应假设，则量 $P(Y|R=r,Z=z;\,\sigma_X)$ 不可识别。在假设 $s=1$ 和 $s'=1$ 成立的情况下，量 $P(Y|R,Z;\,\sigma_X)$ 可识别，且其表达式为



(47)

我们注意到，除了由标量带来的限定之外，这恰好就是 Correa 等人 *[37]* 得出的表达式。不过，即使假设 $s=1$ 成立，若出现 $s'=0$ 的情况（如上所述，这是可计算的），那么显然式 47 并不成立，因此本文的讨论无法得出结论。

虽然这些图示在纸上可能比传统的治疗方案占用更多空间，但在查看图示时，很容易看出这些例子所需的推理步骤，而且这种表示实际上有助于揭示其中的微妙之处。

8 反事实

如果马丁·路德·金在1958年遭刺时就去世了，会怎样？ *[92]*今天，Mary醒来时头痛，但如果她昨晚服用了阿司匹林，今天醒来时还会头痛吗？ *[41]*

上述例子这类反事实问题十分常见，无论是在日常生活、法律语境还是科学领域。比如，问 Mary 如果服用了阿司匹林，是否就不会头痛，隐含的意思是：其他条件都不变，并且已知她实际上没有服用阿司匹林，而且确实出现了头痛。然而，我们实际上无法回到过去，给那个没有服用阿司匹林的 Mary 服用一片阿司匹林，这正是为反事实提供经验基础的根本困难所在

。尽管反事实在人类推理中十分常见，但围绕其方方面面——语义、认知和形而上学 [93]——却引发了诸多争议和哲学辩论。

这里我们将暂且搁置这些问题，只关注因果模型框架 [94, 95, 96, 4, 80, 41, 14]所提供的反事实形式化。这包括对反事实概念和变量的定义，并在此基础上建立一套用于对反事实陈述进行概率推理的“逻辑”，或规则体系。这种形式化依赖于因果模型框架的两个特征。

1. 干预使我们能够精确地设想一个在某种程度上有所改变的假想世界，尤其是将某个变量“固定”为不同于实际发生值的某个值时；2. 函数因果模型中内生变量之间的确定性关系，使我们能够精确界定“其他一切条件不变”的含义。

接下来，我们将以图示方式对这种形式化作基本呈现；事实证明，这是一种非常自然的语言，可用于定义和澄清反事实概念。

8.1 反事实的概念

8.1.1 入门示例

我们先再次考虑这个示例问题：“今天 Mary 醒来时头痛，但如果她昨晚服用了阿司匹林，今天醒来时还会头痛吗？”（另见参考文献 [41]）。令 $X \in \{1, 0\}$ 表示 Mary 是否服用阿司匹林，$Y \in \{1, 0\}$ 表示她是否会头痛。这个问题隐含地指出，在实际世界中，Mary 没有服用阿司匹林（$X = 0$），并且今天早上头痛（$Y = 1$）。

要回答这个问题，必须作出一些假设。令 $\mathbb{M}$ 为图12a所示的结构因果模型，其中包含函数机制 f_X 和 f_Y，以及内生变量之间的因果结构 $X \to Y$；我们假设该模型正确描述了因果依赖关系。首先，考虑这样一种情形：我们还知道，导致 $X = 0$ 和 $Y = 1$的背景变量的精确取值分别为 $U_X = a$ 和 $U_Y = b$，也就是说，图12b中的方程成立。

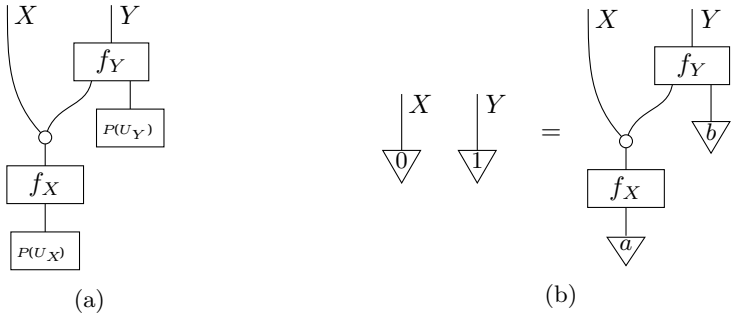
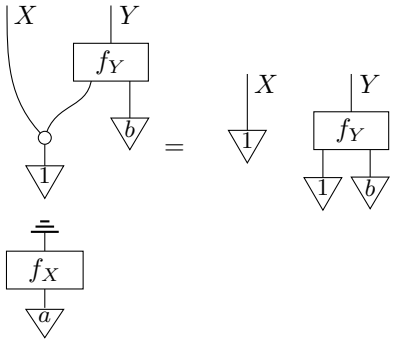


图12

如果我们对因果机制 f_X、f_Y 以及实际世界中外生变量的取值 a 和 b 都有如此完备的了解，那么，Mary 如果服用了阿司匹林是否会头痛这个问题就有了明确的含义和答案。式（48）概括了这一点：等式左侧显示的是与前面相同的模型，其中状态 a 和 b 与实际世界中的状态相同，但另外通过 do-干预将 $X = 1$ 固定下来，从而形式化了“如果她服用了阿司匹林”；随后沿着“1”状态进行复制

确立　　如右侧所示，问题的答案为 $f_Y(1, b$　　$)$.



(48)

然而，即使我们掌握了正确的结构因果模型——知道 f_X、 f_Y、 P（U_X）和 P（U_Y）——通常也无法观测到外生变量的具体取值 a 和 b。不过，实际世界中 $X = 0$ 和 $Y = 1$ 的取值会限制 U_X 和 U_Y 可能的取值。因此，为了回答这个反事实问题——假定 $\mathbb{M}$，但不知道 a 和 b——采用以下约定是很自然的。

首先，我们是在推理两个世界。一个是实际世界，另一个是反事实世界，在这个世界中， $X = 1$ 是人为固定的。无论 U_X 和 U_Y 取何值，它们在两个世界中都应相同。图13a中的图表示通过这种方式将两个世界经由 U_X 和 U_Y 拼接在一起，其中反事实世界中的变量 X 和 Y（的副本）分别记作 X^* 和 Y^*，以作区分。这就是“其他条件均相同”的形式化。

其次，只需在实际世界中对 $X = 0$ 和 $Y = 1$ 进行条件化，即可得出实际世界中的事实如何约束 U_X 和 U_Y 可能取过的值——见图13b。这种条件化使推断影响从实际世界中的 X 和 Y 传递到反事实世界中与之对应的 X^* 和 Y^*，并通过 U_X 和 U_Y 起中介作用。通过直接化简得到的图13b右侧分布 $P(Y^*)$, [26] 因此就是在假设$\mathbb{M}$的情况下，对反事实问题的回答。

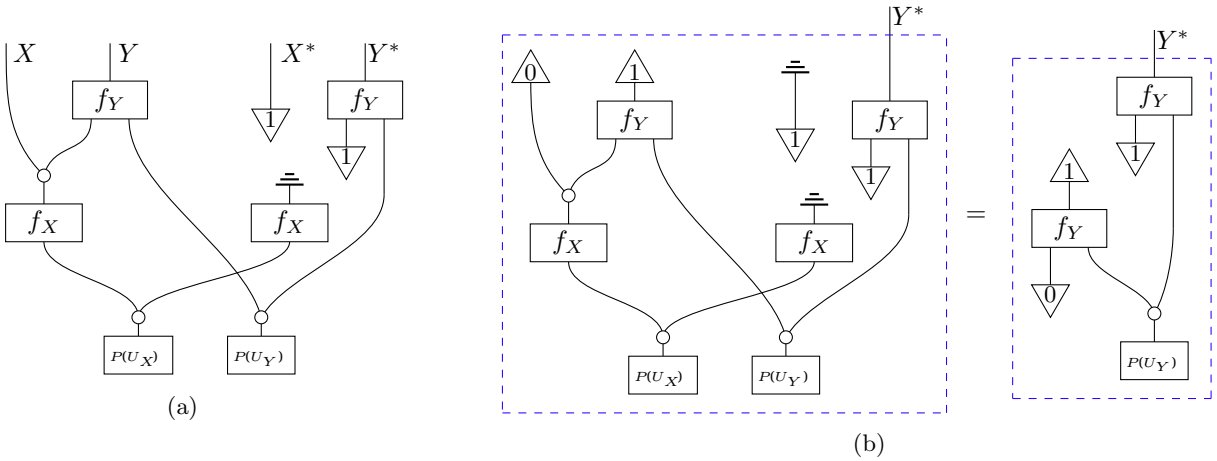


图13

下一节将基于与本例相同的思路，给出反事实的一般定义。这种图解处理与传统文献中的孪生网络方法十分接近（例如，参见文献 [41]）。前述例子的孪生网络图（见图14）展示了实际世界与反事实世界的因果结构如何相互关联。但在这种方法中，贝叶斯更新和计算是分别针对给定数据进行的，而不是像图13b那样全部放在一张图中完成。

[26]请注意，根据假设， $\delta_0 \circ f_X \circ P(U_X) \neq 0$ 成立（毕竟，观测到了 $X = 0$），因此可以根据归一化框的性质省略该标量。

还要注意，后一幅图比孪生网络图包含的信息更多——它明确展示了因果机制的同一性，而孪生网络图只展示了因果依赖的位置。

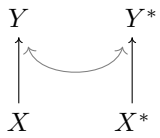


图14

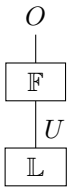
请注意，早在文献 [11] 中，Jacobs 等人就用一个同样简单的例子，说明了以弦图方式处理反事实的方法。除了我们将在第8.1.2节和第8.3节中提出更具一般性的方法之外，另一个区别是，我们的形式化框架也允许以图示方式进行贝叶斯条件化，因此反事实的整个计算过程都包含在图13b中。也就是说，文献中常用来介绍反事实处理方法的所谓「三步法」——溯因、干预和预测（例如，参见文献 [80]）——都可以用一张图来表示。

8.1.2 一般定义

本节给出反事实的一般定义，将上述例子背后的思想形式化。从概念上说，这一定义符合因果模型文献中对该术语的通常理解，包括反事实所涉及的世界不止两个；从形式上说，它的适用范围更广，不局限于 $\mathbf{Mat}_{\mathbb{R}^+}$ 这样的具体范畴。更进一步的概念性推广将留待第 8.3 节讨论。

从形式化定义入手，有助于厘清两件事：一方面是反事实的定义，尤其是明确反事实是相对于哪类数据才有良好定义；另一方面是对认识论问题的研究，例如，给定哪类数据，我们在实践中才能对某个反事实作出明确无歧义的推断。正如第8.1.1节中的例子所示，定义是相对于给定的函数因果模型而言的，因此采用第5节的开放因果模型视角会很有帮助。

以下设 $\mathbf{C}$ 为一个带效应条件化的 cd-范畴（见第2节）。设 $\mathbb{M}$ 是 $\mathbf{C}$ 中的一个函数因果模型，其内生变量为 V，输出变量为 $O \subseteq V$，确定性机制为 $(f_X)_{X\in V}$，外生变量为 $(U_X)_{X\in V}$，相应的状态为 $(\lambda_X)_{X\in V}$（见定义36）。按照例63，令 $\mathbb{F}$ 为相应的确定性部分，其输入为 U，输出为 O，并将其视为一个开放因果模型；再令$\mathbb{L}$ 为 U 上的整体乘积态 $\bigotimes_X \lambda_X$，使得 $\mathbb{M} = \mathbb{F} \circ \mathbb{L}$。

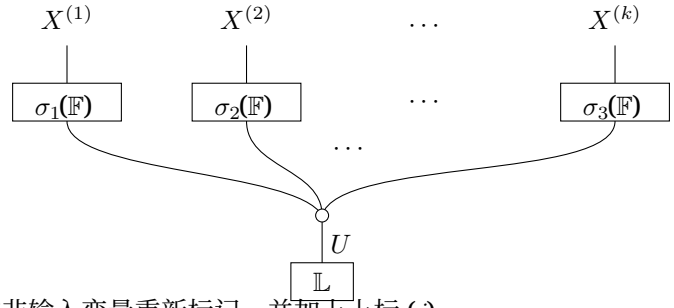


有了这些记号，我们便可以直接给出孪生网络的一般形式，即将两个以上的世界拼接在一起。

定义77.（平行世界模型[27]）：设 $\mathbb{M} = \mathbb{F} \circ \mathbb{L}$ 是 $\mathbf{C}$ 中一个输出变量为 O 的函数因果模型。对于每个 $j = 1, \ldots\ldots, k$，设 σ_j 是对 $\mathbb{F}$ 的干预，使得 $\sigma_j(\mathbb{F})$ 仍然是确定性的。由此诱导的因果模型 $\mathbb{M}^k_{\sigma_j}$:
$= \left(\bigotimes_j^U \sigma_j(\mathbb{F})\right) \circ \mathbb{L}$，即具有下方网络图的模型，称为平行

[27]这一术语体现了它与用“平行世界图”处理反事实的相似性（例如参见参考文献 [41, 97]）。

世界模型。[28]

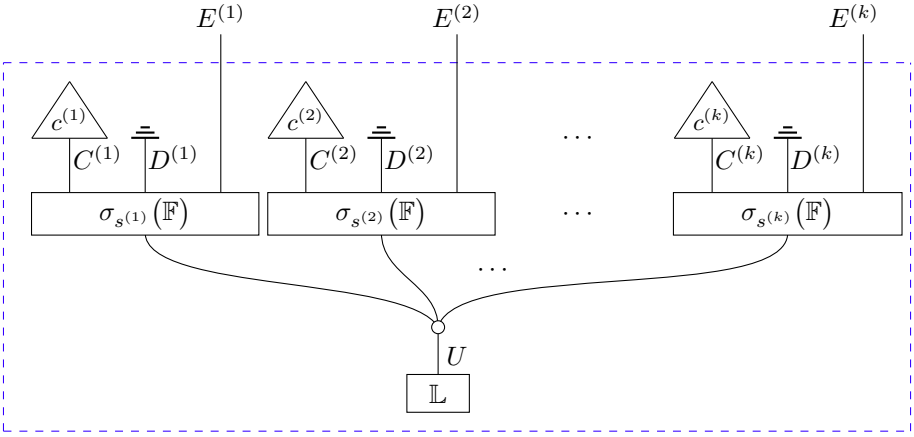


(49)

为作区分，将 $\sigma_j(\mathbb{F})$ 的非输入变量重新标记，并加上上标 (j)。

反事实概念涉及此类平行世界，此处具体使用 do-干预以及条件化。

定义 78。（反事实）：设 $\mathbb{M}=\mathbb{F}\circ\mathbb{L}$ 是 C 中的一个函数因果模型，其内生变量为V，输出变量为 $O\subseteq V$。反事实是如下形式的一个状态：



源自一个平行世界模型，其中每个 $\sigma_{s^{(j)}}$ 对 $S^{(j)}\subseteq O$ 的某个尖锐态 $s^{(j)}$ 执行 do-干预 $\mathrm{do}(S^{(j)}=s^{(j)})$。这里，每个 $O^{(j)}$ 被划分为 $C^{(j)}\cup D^{(j)}\cup E^{(j)}$，且每个 $c^{(j)}$ 都是作用于 $C^{(j)}$ 的尖锐效应。此外，我们要求对于某个 j、j'，满足 $j\neq j'$，

$$C^{(j)}\neq\emptyset\neq E^{(j')} \tag{50}$$

换言之，反事实是一个由 do-干预定义的 k 平行世界模型，并与作用于每个 $C^{(j)}$ 的尖锐效应复合；每个 $D^{(j)}$ 都被丢弃，随后再将所有结果归一化。通常，集合 $S^{(j)}$ 中有一个为空，也就是说，其中一个平行世界由未改变的模型 $\mathbb{F}$ 描述，并被视为“实际世界”。为了与传统记法建立联系，同时又不暗示任何特定范畴（例如 $\mathsf{C}=\mathsf{Mat}_{\mathbb{R}^+}$），我们有时也将定义 78 中由反事实定义的状态记作

$$P_{\mathbb{M}_s^k}\left(E^{(1)},\ldots,E^{(k)}\mid C^{(1)}=c^{(1)},\ldots,C^{(k)}=c^{(k)}\right) \tag{51}$$

其中 s 统称为 $s^{(j)}$ 这一族；为进一步简洁起见，我们也只写作 $P_{\mathbb{M}_s^k}(E\mid C=c)$。以下术语有助于我们简洁地指代反事实中的关键数据。

定义79. 给定 C 中的对象集合 O，世界项 (s,c,E) 由 $S\subseteq O$上的尖锐态 s、$C\subseteq O$ 上的尖锐效应 c 以及与 C不相交的子集 E给出。同一对象集合 O 上的世界项有限族 $W^k=\left(s^{(j)},c^{(j)},E^{(j)}\right)_{j=1}^{k}$称为反事实项，若它们满足某个 j,j' 对应的式 (50)，其中$j\neq j'$。

[28]回顾一下 $\bigotimes_j^U$ 的记法，它表示定义 64 中变量 U 的“共享”。

虽然一族世界项中的数据并未涉及函数因果模型（FCM），但这种记号当然是以定义 77 为背景的，上标 (j) 通常用于标记平行世界模型中的世界。因此，反事实由函数因果模型 $\mathbb{M}$ 以及反事实项共同确定。

备注 80。（与传统记法的关系）：在因果模型文献中，讨论反事实时往往会使用诸如 $P(\gamma|\delta)$ 这样的表达式，其中 γ 和 δ 是反事实表达式的合取式；用我们的记法来写，它们是如下形式的取值赋值

$$\begin{aligned}\gamma &= \left(E^{(1)} = e^{(1)} \wedge ... \wedge E^{(k)} = e^{(k)}\right)\\ \delta &= \left(C^{(1)} = c^{(1)} \wedge ... \wedge C^{(k)} = c^{(k)}\right)\end{aligned}$$

也就是说，δ, γ 本质上与 $C = c, E = e$ 是相同的数据，只不过被理解为逻辑表达式。注意：

- 这种传统处理方式要求将反事实视为概率 $P(\gamma|\delta) \in [0, 1]$。因此，$P(\gamma|\delta)$ 可能没有定义，因为 $P(\delta)$ 可能为零 *[41]*。
  相比之下，如定义 78 中那样，将反事实定义为 $E^{(1)}$,..., $E^{(k)}$ 上的状态，在这个意义上总是定义良好的。例如，在 $\mathbf{Mat}_{\mathbb{R}^+}$ 中，归一化框给出条件概率分布；若 $P(\delta)$ 为零，则给出零态。后者不是概率分布，因为它对定义域中的所有取值都赋值为零，因此无法像 $P(\gamma|\delta)$ 那样用单个概率来表达。

  当然，即使 $P_{\mathbb{M}_s^k}\big(E \mid C = c\big)$ 是一个概率分布，它也可能不具有全支撑，即 $P_{\mathbb{M}_s^k}\big(E = e \mid C = c\big)$ 可能对某些 e取值为零。将相应的尖锐效应代入定义 78 中的开放线即可计算这一点；这些开放线至关重要，因为它们位于归一化框之外。

- 在文献中，不同的世界及其变量通常用下标标记，下标指明该世界的 do-干预作用于哪些变量，例如在引言中的例子里，用 Y_x 而不是 Y^*。这种记法会使定义 78 显得杂乱，因此我们用相应的上标表示第 j个世界。

- $P_{\mathbb{M}_s^k}$ 中的下标明确指出了定义反事实所依据的函数模型。

从更具概念性的角度来看，有几点值得关注和讨论，包括将定义 77 和 78 分开，以及反事实项所满足的特定条件 (50)。

反事实关联着在两个方面有所不同的世界；考虑某种替代情境时，与之相对的事实性有两重含义：

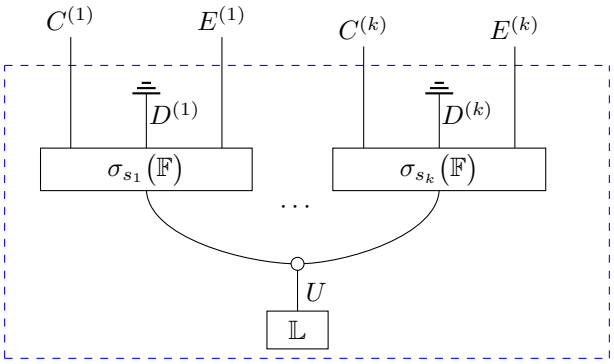
1. 这些世界的因果结构不同。“如果 Mary 吃了阿司匹林，而其他一切都保持不变”这一设想，要求对因果结构进行调整；否则，某个变量 X 的所有确定性成因都相同，就只会产生相同的 X值。以这种方式“介入”，设想某个变量取不同的值，在形式上就是 do-干预，因此它们会出现在定义 78 中。不过，理解这一点的方式并非操作性的，而是对 Lewis 的反事实可能世界语义中与小奇迹相对应之物的形式化 [98, 99].[29]

2. 每个世界都包含其自身的事件类事实——有别于关于因果结构的事实——这些事实由（对）变量 $C^{(j)}$的特定取值 $c^{(j)}$ 进行条件化来刻画。“如果 Mary 吃了阿司匹林”这一设想，只有相对于这样一个事实，即在现实中她没有吃阿司匹林，而且确实头痛，才会使关于她在替代世界中是否头痛的问题成为一个反事实问题。

接下来讨论条件 50。这要求至少在某个世界中，也就是说，对于某个 j，子集$C^{(j)}$ 非空，因此某个世界中的某些变量正处于条件化之下。相比之下，考虑一个类似定义 78 中的状态，但不对效应 $c^{(j)}$进行条件化的状态，也就是表示以下

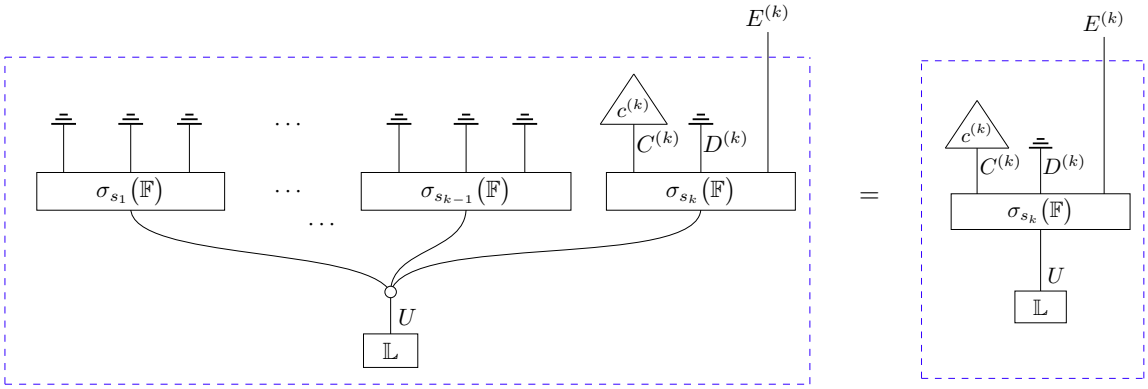
[29]请注意，这里存在 一些关键差异，与 Lewis 对反事实的处理及其更广泛的研究纲领相比尤其如此 e [93, 4].

分布 $P_{\mathbb{M}_s^k}(E, C)$ 在所有 $C^{(j)}, E^{(j)}$ 上的状态：



这些世界的输出变量会相关，因为它们通过背景变量 U连接在一起。然而，根据构造，无论某些特定变量 S_j 是否在某个第 j 个世界中被设定为特定值——这正是定义相应世界的要素——都不会影响其他世界中的概率。只有尖锐效应 $c^{(j)}$ 才会使推断影响通过 U从一个世界“传递”到另一个世界；可以说，这正是反事实的全部要义。

这一点在图示中尤为明显。考虑第 k 个世界上的边缘分布，得到的仅仅是该世界相应的干预模型对 $P\left(E^{(k)}|C^{(k)} = c^{(k)}\right.$的预测；$\left.\text{do}(S^{(k)} = s^{(k)})\right)$，这与反事实毫无关系：



毋庸置疑，平行世界状态和概率 $P(\gamma)$是计算反事实的重要组成部分。不过，定义 78 清楚表明，反事实不能被忠实地改述为思考“如果进行干预会怎样”——它远不止于此。[30] 与此同时，定义 78 的涵盖范围很广，它并不要求在所有世界中都进行条件化，也不要求只在实际世界 $\mathbb{F}$中进行条件化。下一节将讨论一个具体例子，它比阿司匹林与头痛的例子更复杂。

8.2 反事实的可识别性

8.2.1 问题陈述与示例

反事实概念是相对于给定的函数因果模型来定义的。然而，有时可以根据数据明确推断出某个特定反事实的身份，而这些数据并未唯一确定某个函数因果模型；当然，在实践中，人们几乎从来没有这样的模型。

[30]文献中有时会把反事实改述成一个“如果.....会怎样”的问题。这种说法容易造成误解，因为“如果我伸手设置 $X = x$，即对X进行do-干预，会怎样？”这样的问题并不一定涉及反事实。原则上，可以通过实验回答这类问题；在某些情况下，只要有适当类型的数据（见第7节），甚至完全不需要进行任何干预。与此相反，反事实主张原则上不可能通过经验加以验证。图示呈现通过世界和事实性这两个概念，精确地阐明了反事实性所依赖的区别。

这引出了反事实可识别性的认识论问题。用传统术语表述（参见注80），问题如下：给定一个具有顶点 O 的祖先有向混合图 G，以及所有 do- 干预分布的集合 $P_*(O) := \{P(Y; \mathrm{do}(X)) \mid X, Y \subseteq O,$ 其中 $\}$ 两两不相交，反事实 $P(\gamma|\delta)$ 是否可识别？请注意，给定的数据包含所有 do-干预分布。[31] 如果根据此定义，某个特定反事实是可识别的，那么相关的干预分布可以通过相应的实验数据获得；否则，问题就归结为这些干预分布的可识别性问题——这体现了因果层级。 [4, 14]

这里我们将用以下术语来讨论这个问题。设 C 是一个带有效应条件化的 cd-范畴， O 是其中一组对象——将它们视为"观测到的"变量。给定一个顶点为O的祖先有向混合图 G、某个集合 $P_*(O)$ 以及关于集合 O 的反事实项 $W^k = \left(s^{(j)}、\ c^{(j)}、\ E^{(j)}\right)_{j=1}^{k}$，问题是：相应的反事实是否

$$P_{\mathbb{M}_s^k}\left(E \mid C = c\right) \tag{52}$$

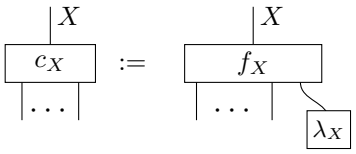
唯一确定，因而可以用 $P_*(O)$ 表示?

鉴于数据并未限定任何特定的函数因果模型 $\mathbb{M}$，在式 (52) 中写出该表达式或许显得奇怪。然而，反事实的形式化依赖于底层的决定论，因此隐含地假定存在这样一个正确的模型。如果某个反事实是可识别的，那么考虑哪个与 G 和 P_* 相容的 $\mathbb{M}$并不重要，因为任何这样的模型都会推出相同的反事实。这与文献中 $P(\gamma|\delta)$ 所起的作用并无不同；在不可识别的情况下，它并不是一个定义良好的量。

对该问题的图解处理将考虑一个图 C，该图引用一个符合给定数据的函数因果模型$\mathbb{M}$，并且可以据此 C 确定反事实是否可识别。因此，无论哪种情况，具体的 $\mathbb{M}$ 都无关紧要——它只是运用图示推理过程中的一个中间步骤。

问题的形式化设定。给定一个顶点为 O的祖先有向混合图 G，考虑其根化结果 $\tilde{\rho}(G)$，其顶点为 $V = O \cup R$，其中 R 是由 $\tilde{\rho}$引入的附加根节点。设 $\mathbb{M} = \mathbb{F} \circ \mathbb{L}$ 是一个相应的函数因果模型，其背景变量为 $\{U_X\}_{X \in V}$，与 P_* 相容，且其因果结构与 G相容，即 $\pi_O(G_{\mathbb{M}}) = G$。随后，我们将构造并考察由该 $\mathbb{M}$ 及反事实项 W^k所诱导的反事实图 C。

作为一种记号约定，对于任意 $X \in V$，我们将根据公式（53），把由二元组（f_X，λ_X）定义的有效因果机制记作 c_X。因此，集合 $\{c_X\}_{X \in V}$ 定义了一个具有有向无环图 $\tilde{\rho}(G)$ 的因果模型。



(53)

注释 81.（根化的作用）：请注意，我们明确指定了第 6.2 节中定义的根化方法 $\tilde{\rho}$。这是因为，对于任意给定的祖先有向混合图 G，另一种根化方式 $\rho(G)$可视为 $\tilde{\rho}(G)$的特例：其中一些具有两个以上子节点的根节点，实际上是由多个根节点组成的集合，而其中每个根节点都只有两个子节点。在第 7.1 节讨论因果效应可识别性时，我们援引了参考文献 *[38],* 中的一个结果，即定理 72，以说明对祖先有向混合图进行何种根化并不重要。然而，对于反事实而言，这一结果并不能直接适用。虽然乍看之下，处理祖先有向混合图中双向边的方式为何不影响反事实的可识别性似乎并不明显，但遗憾的是，诸如 *[41]* 之类的研究并未探讨这一问题。我们也将把这个问题留待以后讨论，但为求稳妥，采用 $\tilde{\rho}$。[32] 最后我们指出，对于 *[41],* 中的任何例子（其中许多将在下文讨论），$\tilde{\rho}$ 和 ρ 无论如何都会得到相同的有向无环图。

[31]请注意，对于 $X = \emptyset$，$P(O)$ 包含在集合 $P_*(O)$ 中。[32]这意味着，在可识别性的情况下，这是尽可能强的主张；而在不可识别性的情况下，如果假设所有双向边都对应于具有两个子节点的根节点（正如根化方法 ρ 所设定的那样），那么这本身并不意味着不可识别性。我们将在注释 93 中回到这个问题。

正如因果效应的可识别性一样，反事实的可识别性问题也已经得到解决——存在一种对此问题可靠且完备的算法 [41]。后续章节将讨论该解决方案的图解处理。

不过，在此之前，考察文献中的一个例子，并在我们的图示框架中处理它，会很有启发性。这有几个目的。首先，它阐释了我们在定义 78 中给出的、较为抽象的反事实定义。其次，它为后续的一般算法解法建立必要的直觉。第三，我们能够讨论这个例子，并以直接明了的方式给出解法，而无须证明任何关于反事实的事实，只需借助图示推理——运用范畴中已有的重写规则——这一点本身就足以说明问题。

例 82。Shpitser 和 Pearl 在文献 *[41]* 中研究了以下例子。考虑图 15a 中的祖先有向混合图 G，并假设给定 $O = \{X, Y, Z, W, D\}$ 时， $P_*(O)$ 已知。反事实 $P_{\mathbb{M}^3_s}\big(Y_{(1)} \mid X^{(2)} = \tilde{x}, D^{(2)} = d, Z^{(3)} = z\big)$（其中，定义世界的 s 表示 $\mathrm{do}(X^{(1)} = x)$ 和 $\mathrm{do}(D^{(3)} = d)$[33] ）是否可识别？他们还描述了一个医疗情境，由此提出了这个涉及三个世界的查询：“例如，我们可能想知道，在已知患者服用了剂量为 $\tilde{x}$ 的药物 X、剂量为 d 的药物 D，并且知道中间症状 Z 对治疗 d 的反应的情况下，患者服用剂量为 x 的药物 X 后出现症状 Y 的可能性有多大。”

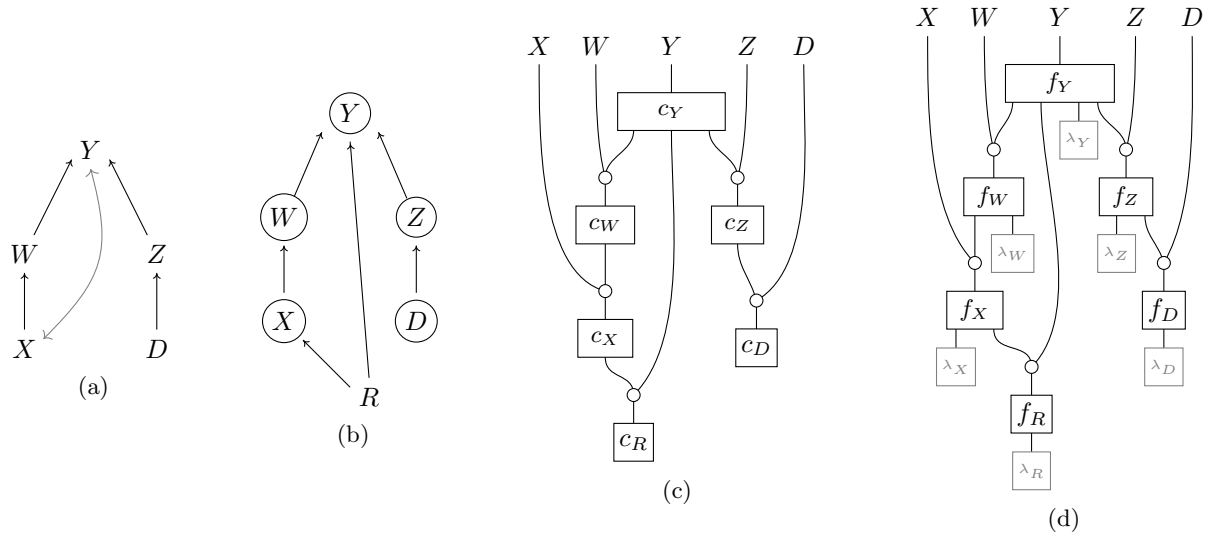
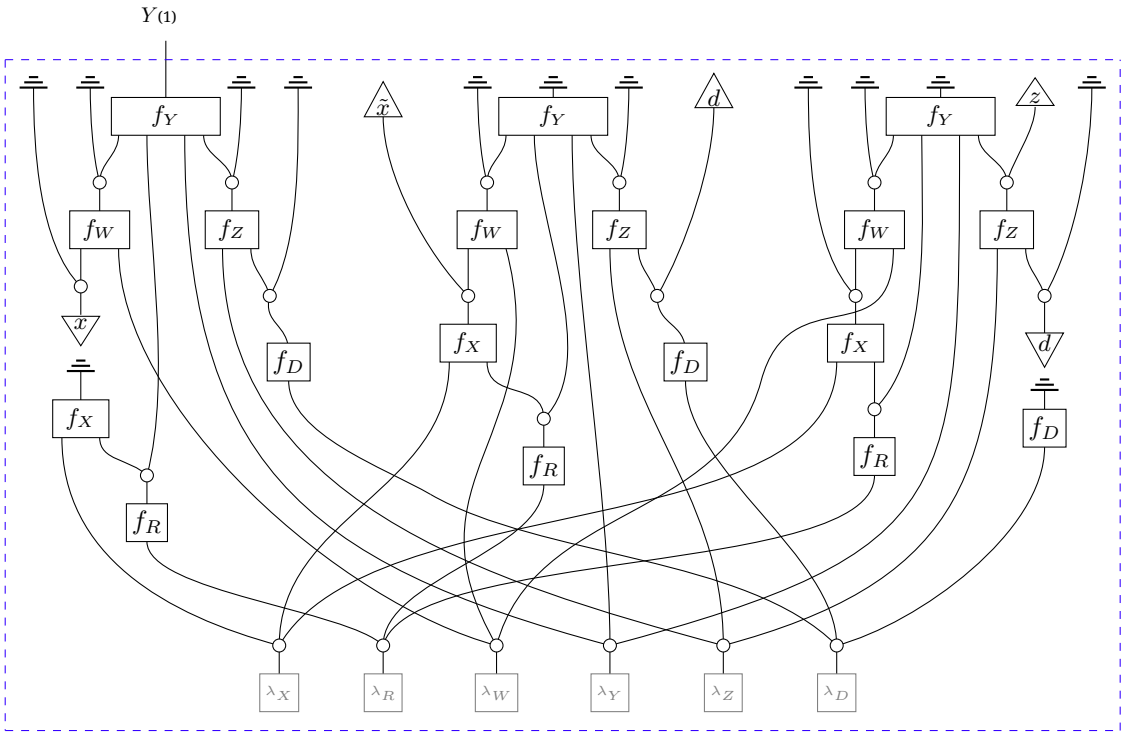


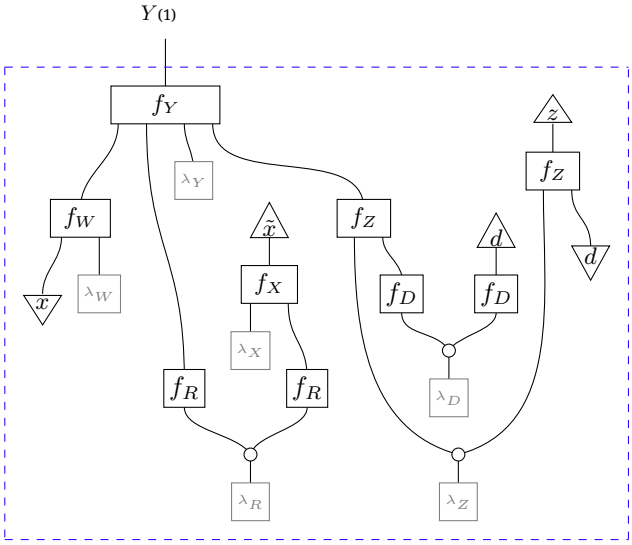
图15

对于图解处理，请考虑图15b中的根化 $\tilde{\rho}(G)$ 以及相应的网络图 $D_{\tilde{\rho}(G)}$ 图15c。图15d则展示了相应的函数因果模型，我们将以此为基础考察感兴趣的反事实。该反事实为

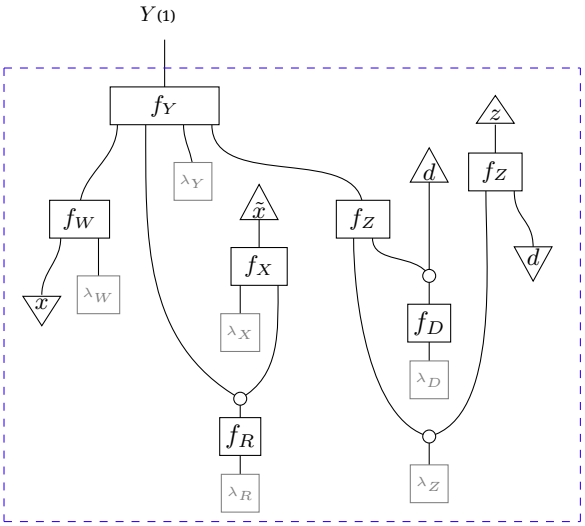
[33]也就是说，反事实项 W^3 与 $S^{(1)} = \{X^{(1)}\}$, $s^{(1)} = x$, $E^{(1)} = \{Y_{(1)}\}$, $C^{(1)} = \emptyset$, $S^{(2)} = \emptyset$, $E^{(2)} = \emptyset$, $C^{(2)} = \{X^{(2)}, D^{(2)}\}$ $c^{(2)} = \tilde{x} \otimes d$ $S^{(3)} = \{D^{(3)}\}$ $s^{(3)} = d$ $E^{(3)} = \emptyset$ $C^{(3)} = \{Z^{(3)}\}$ $c^{(3)} = z$,,,,,,.

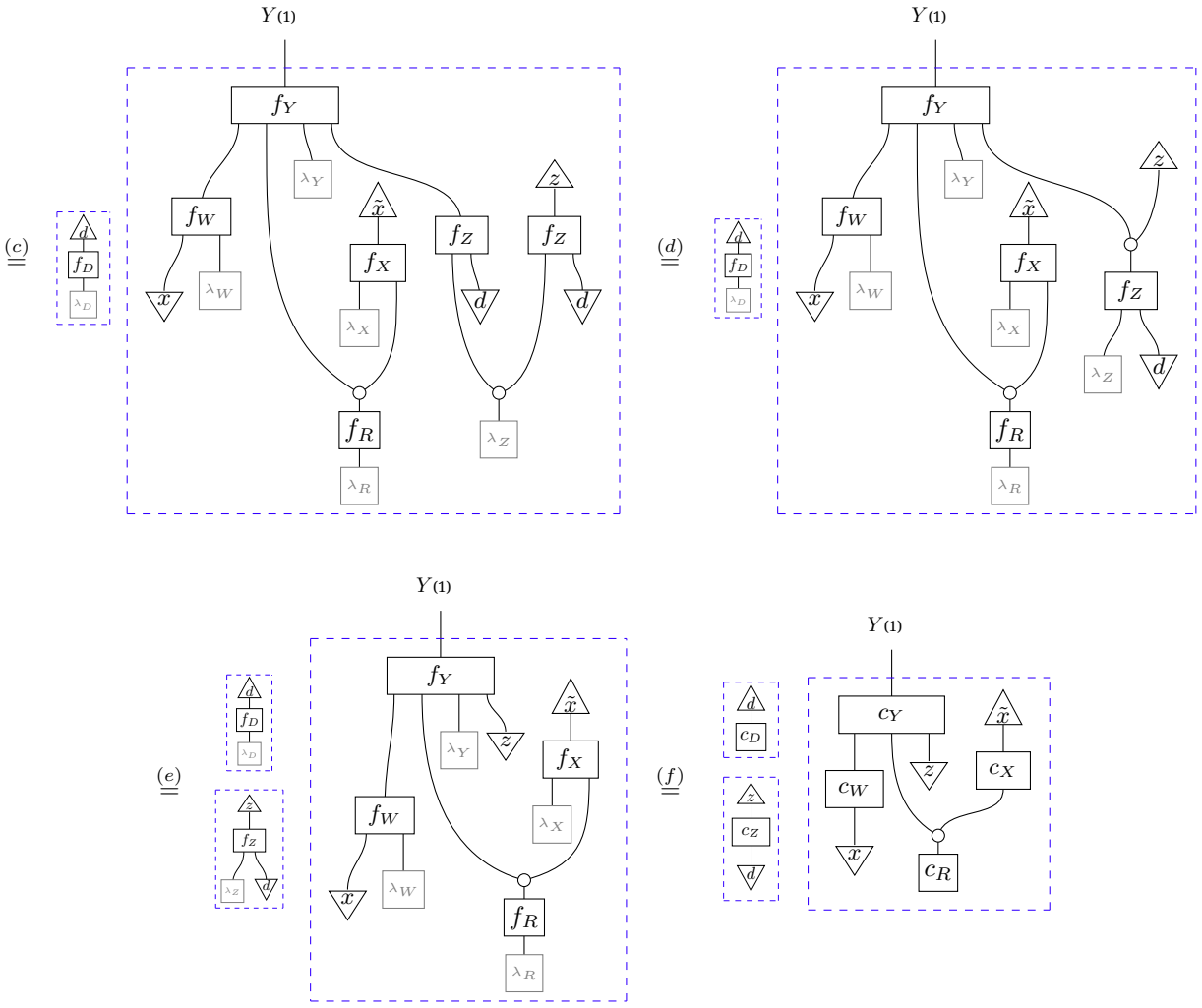


(a) $\equiv$



(b) $\equiv$





步骤（a）只是利用了丢弃操作在连接到复制映射时会“贯穿”机制并移除导线这一性质，正如第7.3节中所用。步骤（b）利用了 f_D 是确定性的这一点，因此可以将 f_D 的两个副本通过复制映射移动，从而留下一个输出被复制的 f_D；函数 f_R也同理。步骤（c）利用了确定性效应 d与复制映射相连这一点，因此可将 $d \circ f_D \circ \lambda_D$分离出来作为标量（见第2.4节）。步骤（$d$）和（$e$）重复了类似的操作——首先将 f_Z 向下通过复制映射移动，然后利用效应 z 与复制映射相连这一点。最后，步骤（f）只是根据图15c与图15d之间的等价关系，将函数和背景噪声分别替换为对应的信道。

请注意，这与标量在上面的图示中浮动，ft十分重要——相应的归一化标量要么为一，要么为零；后一种情况下，反事实的整个状态都会求值为零态（见注80）。还要注意，虽然属于不同世界的线可能标有不同的标签，例如示例中的 $Z^{(1)}$ 和 $Z^{(3)}$，但它们在范畴中对应同一个底层对象，因此像 f_Z 所示的函数“穿过复制”是定义良好的。为避免杂乱，大多数线的标签都省略了；但凡需要追踪变量所属的世界——因为我们要计算的反事实会引用这个特定变量——该变量要么是开放线且仍带有标签，要么其所依据的取值会指明所属的世界。因此，不存在歧义的风险。

现在，根据 观察图15c中的网络图，可得出以下等价关系（显而易见 ）图16。

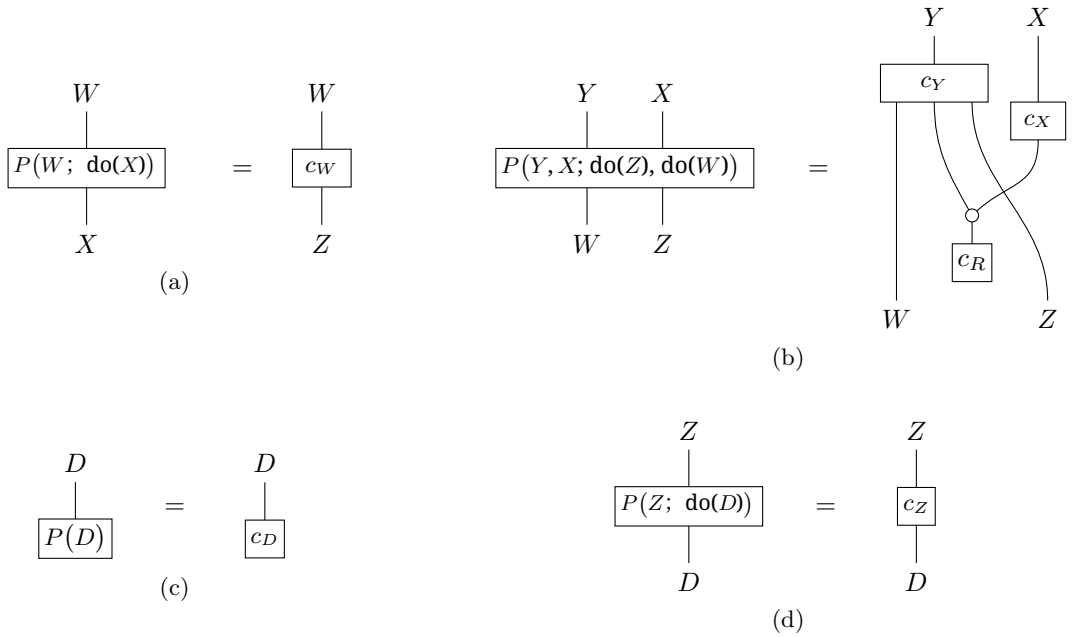
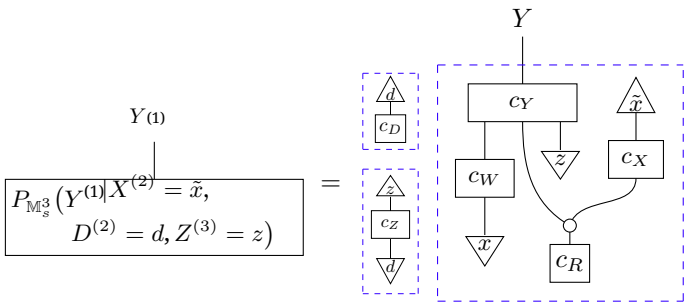


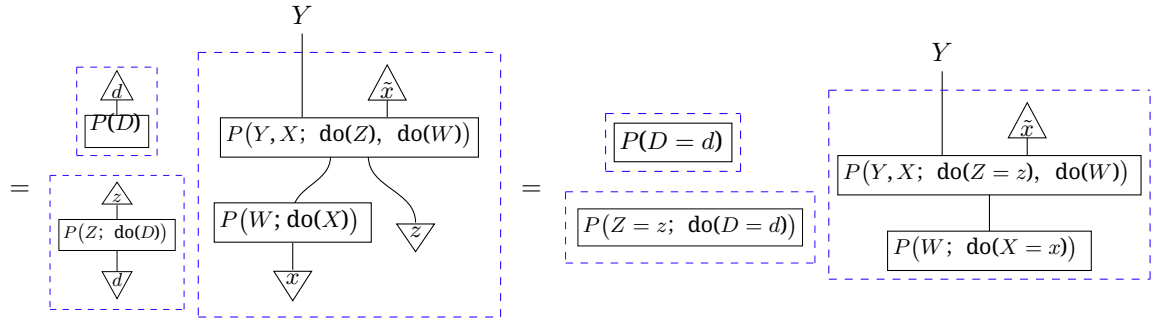
图16

总而言之，我们发现：

$$P_{\mathbb{M}_s^3}\left(Y^{(1)} \middle| X^{(2)} = \tilde{x},\ D^{(2)} = d, Z^{(3)} = z\right) = \tag{54}$$



$$= \quad = \tag{55}$$



由于公式 55 的右侧完全由可识别表达式构成，我们得出结论：这个反事实具有可识别性。请注意，这实际上与文献 *[41],* 中给出的解相同，只是这里的标量还明确说明了该反事实取值为零态的条件。

8.2.2 化简的一般方法——simplify-cf算法

上一节对例 82 的讨论分为两部分。首先，步骤 (a) 到 (f) 应用了各种重写规则，将反事实图转换为一个等价但简单得多的图。其次，简化图为研究给定 P_* 时该反事实的可识别性提供了依据。这种两步方法与文献中对反事实的传统处理方式直接对应。为了更明确地说明这种对应关系，尤其是为了以图示术语给出一般性方案，本节提出以下算法，用以概括例 82 中所用的这类图示简化方法。

函数 simplify-cf(D, V, π):

输入：范畴 C 中的图 D，由一个平行世界模型的网络图（可能与某些效应复合而成）、该平行世界模型所依据的函数模型 $\mathbb{M}$ 的变量集合 V，以及 DAG $G_{\mathbb{M}}$ 上 V 的一个拓扑排序 [34] 构成。

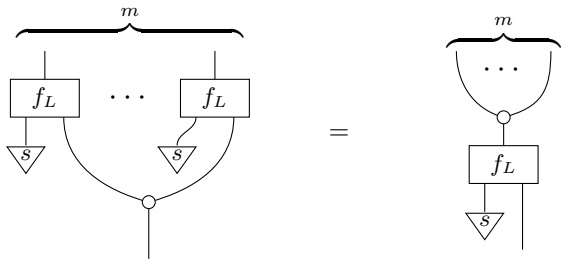
输出：化简后但与原图等价的图 D。

1. 让所有丢弃操作“贯穿”，即反复使用信道的定义性质（定义3），并在被丢弃的导线与复制映射相连时将其删去（定义2），直到图中不再有丢弃操作。

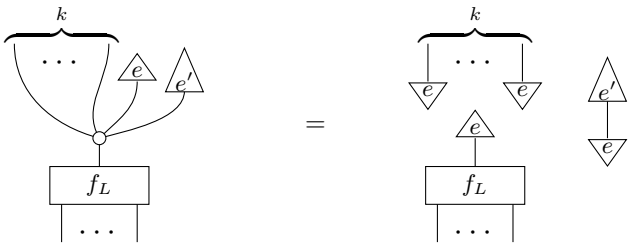
2. 凡尖锐效应与复制映射相连之处，使用等式（6）将所有涉及的导线彼此“分离”。

3. 从最低的根节点开始，按照顺序 π 依次遍历变量 $L \in V$，并对相应变量 L 应用以下步骤。

   (α) 考虑来自不同世界的函数机制 f_L 的所有此类出现 m：它们的输入以如下左侧所示的方式共享，并据此重写 D：



   (β) 如果尖锐效应随后通过复制映射连接到 f_L 的输出，则重写为：



4. 对于任何 $X \in V$，若 f_X 以状态 λ_X 直接输入的形式出现在 D 中，也就是说，并未通过复制映射与 f_X 的其他出现共享该状态，则按照公式 (53) 将其替换为 c_X。

5. 输出重写后的图 D。

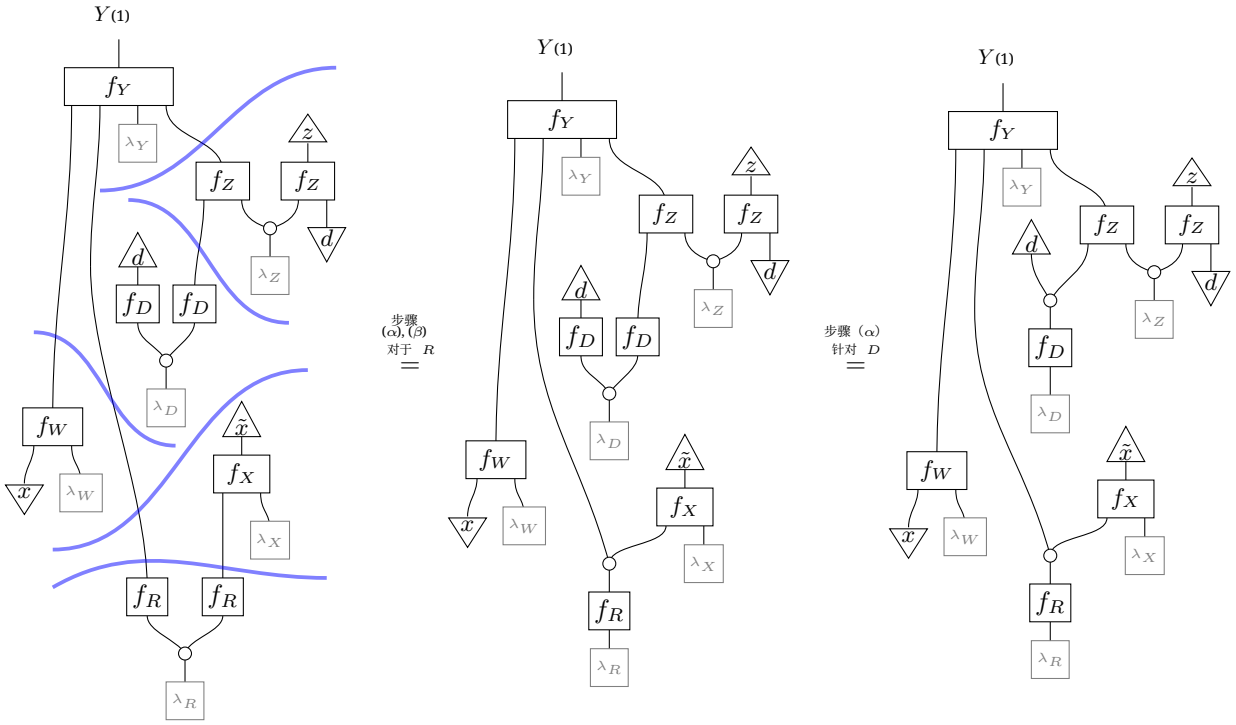
该算法以此类图作为输入并对其进行简化。这类图源自反事实，不过不包含归一化框，因为这些简化都与归一化无关。

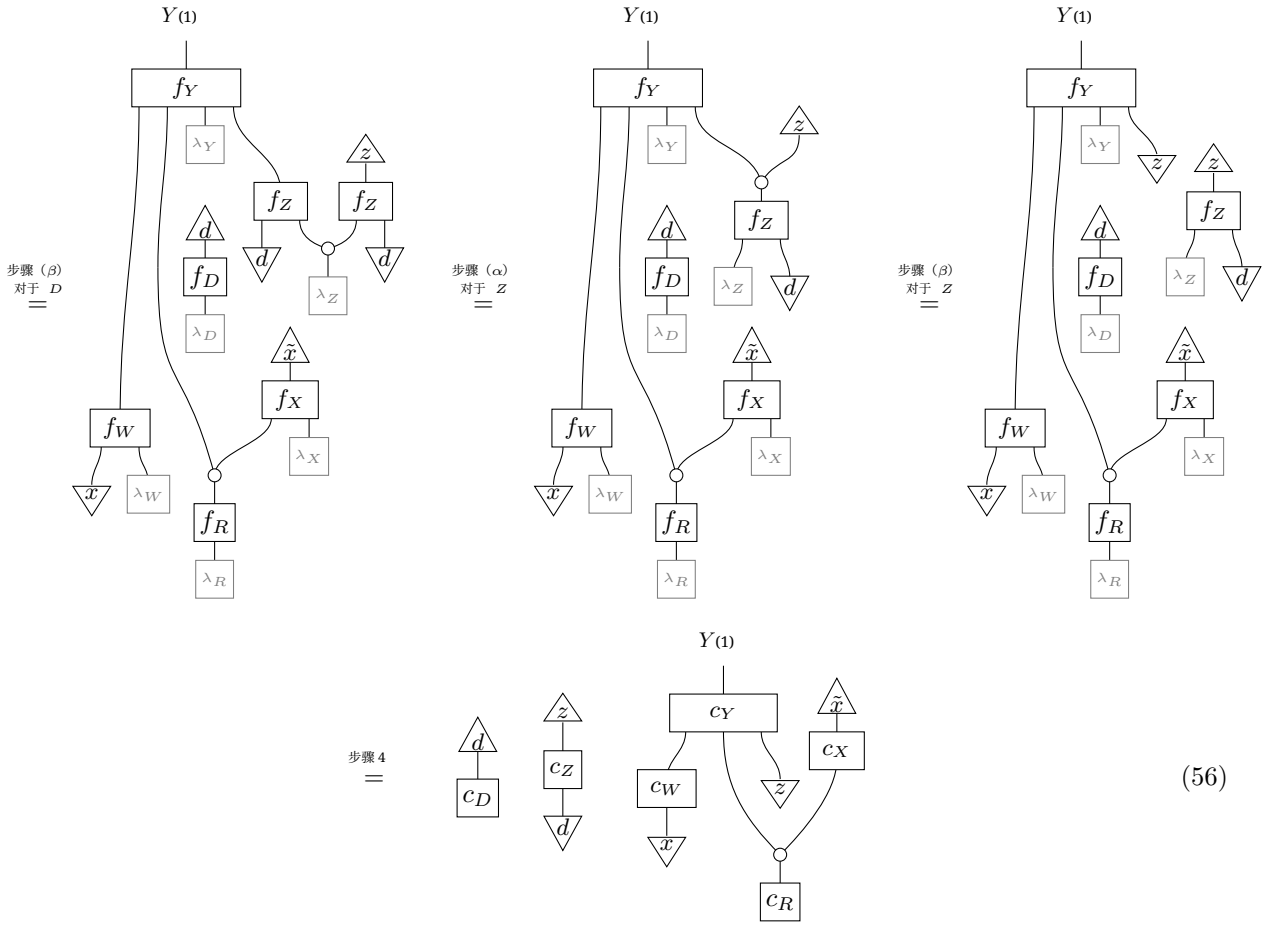
[34] 给定一个有向无环图 G，其顶点为 V，*topological order* 在 V 上是一个与 G 相容的全序，也就是说，若 X 是 G 中 Y 的祖先，则 $X < Y$。

因此，对于反事实项 $W^k = \left(s^{(j)}、\ c^{(j)}、\ E^{(j)}\right)_{j=1}^{k}$ 对应的反事实图 $C = \mathrm{norm}(D)$，其中 D 是不含归一化的对应图，算法重写的是 D 。请注意，在 $\mathbf{Mat}_{\mathbb{R}^+}$ 中， D 表示状态 $P_{\mathbb{M}_s^k}\left(E^{(1)}、...、\ E^{(k)}、\ C^{(1)} = c^{(1)}、...、\ C^{(k)} = c^{(k)}\right)$，它还不是关于 $E^{(1)}、...、\ E^{(k)}$ 的正确归一化概率分布，而是一个函数，对 E 的每个取值赋值都返回相应的概率。反事实 C 则是对应的归一化条件分布。

该算法的一个关键思想是按照拓扑序 π进行重写：起初，对于某个 L ， f_L 的不同出现形式可能并非都已将其所有输入连接到复制映射；但根据 π，随着不断通过函数进行复制，这种情况可能会发生，复制映射也因此“向上移动”。该算法是可靠的，也就是说，输出确实等价于输入 D；这一点显而易见，因为它只使用了在 cd-范畴 C 中普遍成立的等价关系。

为说明该算法，我们再次考虑例 82。该例中重写的步骤 (a) 是 simplify-cf 的步骤 1 和 2 的结果。下方左侧重现了这些步骤的结果，并按照相应蓝色切片所示的拓扑序 R, X, W, D, Z, Y 重新排列。随后根据 simplify-cf 的步骤 3 对其进行化简，其中 (α) 和 (β) 对 X, W 和 Y 均不起作用，而 (β) 对 R 不起作用。





(56)

显然，在整个输出态外添加一个归一化框，得到的恰好是例 82 中 (f) 右侧的表达式。

请注意，simplify-cf 的步骤 4 通常并不会消除背景变量 U_X 的状态 λ_X 的所有出现，这些状态将不同的世界连接起来——这一事实与可识别性判据密切相关，并将在下一节中进一步讨论。

注 8　　3.（关于 simplify-cf 的评论）：关于 simplify-cf 的一些其他有用观察　　:

1. 使用步骤 $3(\alpha)$ 和 $3(\beta)$ 中的改写规则时，不会对变量所属的世界产生歧义——感兴趣的反事实所引用的变量，要么是开放线，即 $E^{(1)},...,E^{(k)}$，因此仍带有其标签；要么是被条件化的变量，此时对应的取值会指明它所属的世界，即 $c^{(1)},...,c^{(k)}$。2. 在步骤 $3(\alpha)$ 中，状态 s 的导线和开放输入线可以是复合的，也就是说，它们可以代表一组变量；此时，它们分别被理解为相应尖锐态的积和复制映射的积。3. `simplify-cf`的输出图由标量的积和一个（积）状态构成；由于由复制映射和单输出通道组成的网络图具有特殊结构，输出中不会出现丢弃操作。

≒

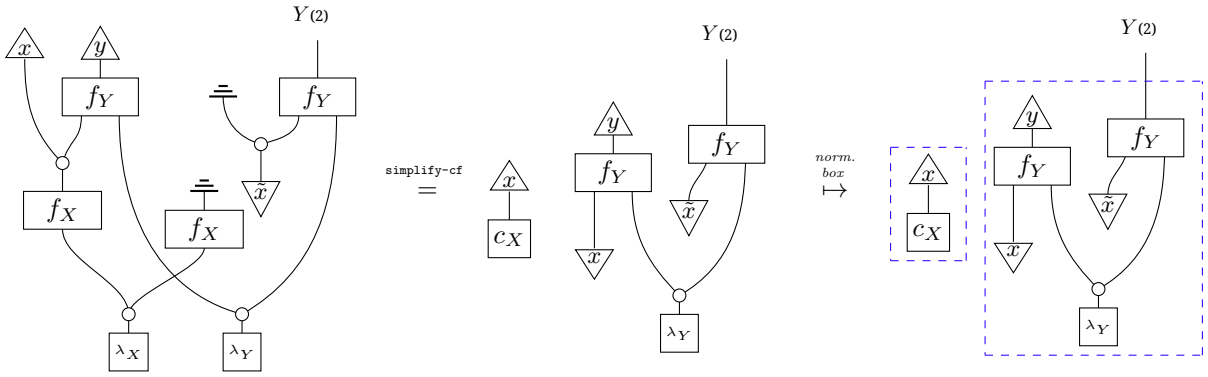
注84。（与make-cg算法的关系）：回顾注80，给定一个祖先有向混合图 G 以及表达式 γ、δ，Shpitser和Pearl提出的算法 *IDC** $(G、\gamma、\delta)$ 会输出 $P(\gamma|\delta)$ 是否可识别，以及在可识别的情况下其具体形式。假设 $\gamma \wedge \delta$ 和 δ 并不矛盾——如果矛盾，则会通过我们的归一化框加以处理——那么 *IDC** 的make-cg $(G、\gamma \wedge \delta)$ 子程序会构造一个平行世界图，随后根据所假设的底层函数关系，通过合并并重新标记节点来简化该图，并输出相应的祖先有向混合图 G'，称为反事实图。可识别性尤其是依据以下内容来判断：

G'。不以正式且详细的方式阐明这种对应关系，simplify-cf与之的关系如下：如果 $D_{\tilde{\rho}(G')}$ 是通过对反事实图 G' 进行根化所得到的网络图，那么根据 δ 对不属于 $\gamma \wedge \delta$ 的任意变量进行边缘化，并与尖锐效应 $c^{(1)}$、……、$c^{(k)}$ 复合，所得图与对应反事实图的simplify-cf输出等价。直观地说，不难看出，按照拓扑序对函数 f_L 进行“传递复制”，实现的效果与make-cg算法中作为其一部分的“合并顶点”相同。

8.2.3 识别算法 id-cf

本节旨在陈述一种算法，它推广了例 82 的第二部分；也就是说，该程序先接收经 simplify-cf 简化后的图，再检查是否可以将其重写为关于 P_* 的表达式；如果可以，就实际执行重写。可以将其视为把 Shpitser 和 Pearl 在文献 [41] 中提出的算法解法的核心思想（尤其是可识别性的充要条件）自然地转化为图示术语。为此，下文首先讨论文献 [41] 中两个用于说明不可识别性的教学示例，借此帮助建立对算法中将要用到的可识别性一般条件的直觉。与目前的处理方式一样，该算法将针对任意合适的 cd-范畴 C 陈述；不过，以下两个取自文献的示例假设环境范畴为 $\mathsf{Mat}_{\mathbb{R}^+}$。

示例 85。首先，再次考虑第 8.1.1 节中的阿司匹林与头痛示例。用上标 (1) 表示实际世界，并用 x 和 y 代替 0 和 1，表示该世界中的条件化；用 (2) 表示由 do($X^{(2)} = \tilde{x}$) 定义的反事实世界。$P_{\mathbb{M}^2_s}(Y_{(2)}|X^{(1)} = x, Y_{(1)} = y)$ 的反事实图（与图13b 相同，但经过归一化并重新命名）如下所示，左侧为该图，旁边是 simplify-cf 的输出。



正如参考文献 *[41]* 所述，只要 $x \neq \tilde{x}$，这个反事实就不可识别。看一下简化图，原因显而易见：对于 $x \neq \tilde{x}$，两个方框 f_Y 无法通过复制映射向下移动，因此 simplify-cf 会得到这一结果。P_* 包含 c_X 和 c_Y，但由于 U_Y 未被观测到，P_* 无法让人推断 Y 如何通过 f_Y 依赖于 U_Y，而这正是识别所求反事实所必需的。

上述例子传达的直觉是：每当某个背景变量（如 U_Y）无法像例 82 中那样被“简化掉”，而是仍保留在 simplify-cf 输出的图中时，这就构成了不同世界之间一种“不可约”的相关性，因此无法根据 P_* 识别该反事实。

例 86. 再来看一个例子，考虑图 17a 中的祖先有向混合图 G、图 17b 中的根化 $\tilde{\rho}(G)$，以及图 17c 中对应的函数因果模型 $\mathbb{M}$。考虑状态 D，它表示 $P_{\mathbb{M}^3_s}\big(W_{(1)\,1}、W_{(1)\,2}、Y_{(2)}\,Z^{(3)}\big)$，该状态源自一个基于 $\mathbb{M}$ 的平行世界模型，以及世界项 W^3；其中世界 (1) 是原始世界，即 $S_1 = \emptyset$，而世界 (2) 和 (3) 分别由 $S_2 = \{X\} = S_3$、$s_2 = x$ 和 $s_3 = x'$ 定义。注意，这些世界项不包含任何条件化——它们不是反事实项，但通过选择适当的尖锐效应即可使其成为反事实项。

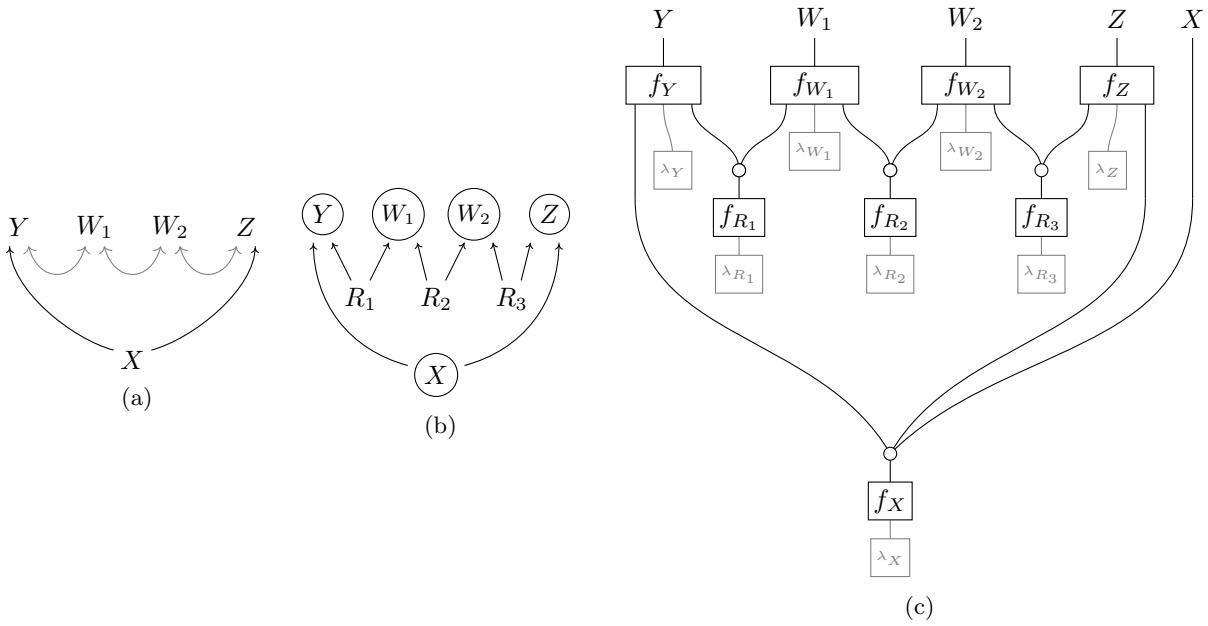


图17

simplify-cf 输出的简化图 D 如图 18a 所示。正如参考文献 *[41]* 中所论证的，分布 $P_{\mathbb{M}_s^3}(W(1)\,1\,, W(1)\,2\,, Y(2)\,Z^{(3)})$ 无法根据 G 和 P_* 识别。请注意，这一次，对于 $\tilde{\rho}(G)$ 中的每个顶点 A （及其关联变量），simplify-cf 都将 λ_A 吸收到相应的信道 c_A 中。尽管如此，这仍是一个不可识别情形。为直观说明这一点，图 18a 将根化 $\tilde{\rho}(G)$ 中引入的所有对应于“未观测”变量 R_1、R_2、R_3 的导线标为灰色。注意，集合 $\{Y, W_1, W_2, Z\}$ 在 G 中构成一个 c- 分量（见第 7.2 节），其中 R_1、R_2、R_3 是显式的潜在结构。虽然图 18b 所示的 $P(Y, W_1, W_2, Z; \mathrm{do}(X))$ 按定义属于 P_*，但图 18a 中的状态并不属于其中，且无法由其识别，因为 c_Y 和 c_Z 接收的是不同的状态 x 和 x'。若要识别所求状态，就需要分别获得 c_Y 和 c_Z，但这与 R_1 和 R_3 未被观测到相矛盾。

对于由 D 得出的任何反事实，只要额外对变量 $W_1^{(1)}, W_{(1)\,2}, Y_{(2)}\,Z^{(3)}$ 中的某些变量进行条件化，同样的不可识别性结论便立即适用。

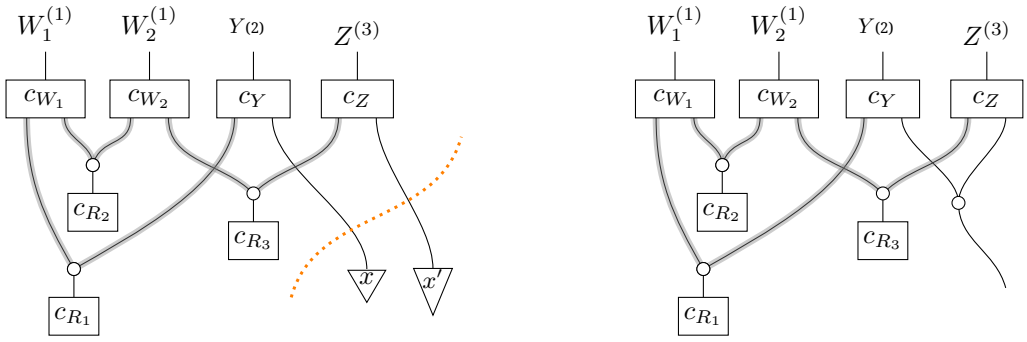


图18：(a) 展示了 simplify-cf 对 $P_{\mathbb{M}_s^3}\left(W_{(1)\,1}, W_{(1)\,2}, Y_{(2)}\,Z^{(3)}\right)$ 的输出 D。橙色虚线勾勒出 D 的一个 R-片段（见下文定义87）；(b) 展示了 $P(Y, W_1, W_2, Z; \mathrm{do}(X))$。

前一个示例　　这个例子表明，引入一个概念来指代网络图中的某个部分会很有帮助，
与之对应的 rk 图对应于祖先有向混合图中的一条混杂路径。

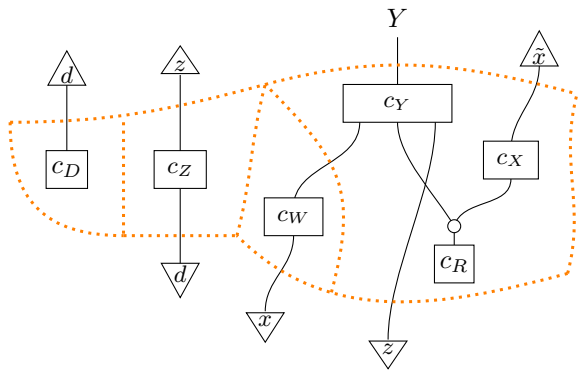
定义 87. 给定一个顶点为 O 的祖先有向混合图 G，令 R 为根化 $\tilde{\rho}(G)$ 引入的附加根节点集合，并令 $D_{\tilde{\rho}(G)}$ 为 C 中相应的网络图。设 D 是一个字符串图，仅由出现在 $D_{\tilde{\rho}(G)}$ 中的机制、复制映射，以及对象 O 的状态和效应组成。[35] 那么， D 的一个 R-片段是 D 的一个极大片段，它仅由那些机制$\{c_X\}_{X\in S_O\cup S_R}$（以及复制映射）组成，其中某个 $S_O \subseteq O$ 和 $S_R \subseteq R$ 满足

- 在 G 中的 S_O 的任意一对顶点之间，都存在一条混杂路径[36]，且该路径上的所有顶点都位于S_O 内部，而路径上的双向边对应于 $\tilde{\rho}$（G）的根节点，并包含在 S_R 中；

  S_O 中这些机制与 S_R 中那些机制之间的连线方式如 $D_{\tilde{\rho}\ (G)}$ 所示（但不对机制 c_X 的其他输入和输出作出任何条件限制 $X \in S_O$）。

示例 88。从沿橙色虚线切割后得到的图18a中的图片段，是相对于图 17a 中的祖先有向混合图 G 的该图的一个 R-片段，其中 $R = \{R_1, R_2, R_3\}$ 如图 17b 所示。

示例 89。再次考虑例 82。对于本例中的单个附加根节点 $R = \{R\}$，由式 56 中 simplify-cf 的输出得到的四个R-片段（由橙色虚线切分）如下:



(57)

鉴于前面已经讨论过的例子，现在不难提出一种算法，将 Shpitser 等人算法解法中的相同思路融入我们的图解处理中。 [41] 下面我们给出该算法可靠性的证明（概要），不过值得指出的是，这个证明几乎是显而易见的——该算法不需要任何非显然的事实或操作，因此无需另行证明；它完全依赖于带效应条件化的 cd- 范畴中的重写规则。

函数 id-cf$(G, W^k, P_*(O))$:

输入：带有顶点 O 的祖先有向混合图 G（其中顶点属于 C）、反事实项 $W^k = \left(s^{(j)}, c^{(j)}, E^{(j)}\right)_{j=1}^{k}$ 以及集合 $P_*(O)$。

输出：失败，或与 W^k对应的反事实 C（假设 G），并用 $P_*(O)$ 表示。

1. 令 R 为根化 $\tilde{\rho}(G)$ 引入的附加根节点集合。令 $\mathbb{M} = \mathbb{F} \circ \mathbb{L}$ 为相应的函数因果模型，其内生变量为 $V := O \cup R$，背景变量为 $\{U_X\}_{X\in V}$，满足 $\pi_O(G_{\mathbb{M}}) = G$，且与 P_*兼容。对于任意 $X \in V$，记 c_X 为将 λ_X 输入 f_X 后按式 (53) 得到的相应概率机制。最后，令 C 为基于 $\mathbb{M}$ 由 W^k 定义的反事实，令 D 为归一化意义下相同的图，C =范数(D)。

2. 对于 $\tilde{\rho}(G)$ 的某个拓扑序 π， D =simplify-cf(D, V, π)。

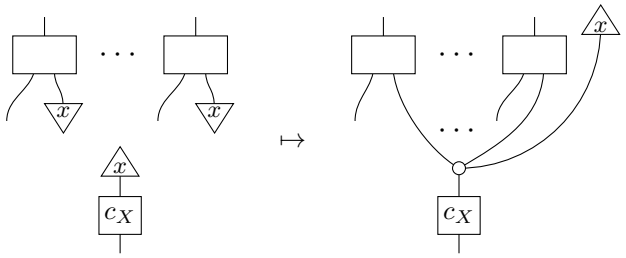
3. 如果 λ_X 未被 simplify-cf 吸收到 c_X 中，则输出 FAIL；否则继续。

4. 对于 $F_l \in \{F_1, ..., F_m\}$， D 的 R片段集合:

[35]机制的相等性是指它们作为 C 中的态射相等；只要导线指向相同的对象，其标记方式并不重要。[36]回顾第 6.1 节，祖先有向混合图中的混杂路径是仅由双向边构成的路径。

1. 对于每个 $X \in O$，使得 X 作为一条导线的类型出现在 F_l 中：

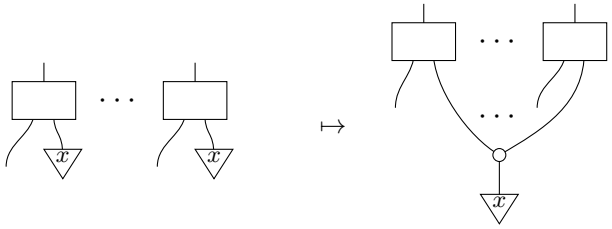
1。如果机制 c_X 是 F_l的一个组件：a. 如果 c_X 的输出导线上没有任何尖锐效应 x 作用于其上，也就是说，它被输入到另一个R片段中，或是 D的一个输出，并且 F_l中没有出现其他类型为 X 的导线，则不做任何操作。b. 如果 c_X 的输出导线在 D 中与某个尖锐效应 x 组合，并且 F_l 中所有其他类型为 X 的导线都输入了状态 x，则按以下方式重写 D：



(58)

c. 否则输出失败。

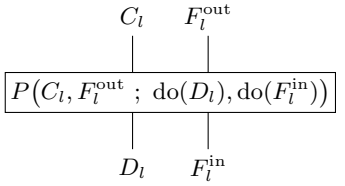
2. 如果机制 c_X 不是 F_l的一个组件：a. 如果所有输入到 F_l的类型为 X 的导线，都通过复制映射连接到另一个 R片段中 c_X 的同一个输出，则不做任何操作。b. 如果所有输入到 F_l的类型为 X 的导线都输入了相同的状态 x，则按以下方式重写 D：



(59)

c. 否则输出失败。

~2. 按以下方式替换重写后的 R片段 F_l ——与 F_l 相同，但多出复制映射——：



(60)

~其中，D_l、F_l^{in}、C_l和 F_l^{out}分别是：输入到 F_l且向其中馈入尖锐态的导线的对象集、F_l 的其余输入、从 F_l 输出且其上带有尖锐效应的导线，以及 F_l的~其余输出。

5. 输出范数(D)。

为了说明这一点，我们再次考虑例 82，尤其是公式57，其中切片标出了 simplify-cf 输出中的 R片段。id-cf 的步骤（3.）和（4.1.）中的条件均已满足，且没有进行任何重写。请注意，我们先前使用的图16a–16d中的等式，恰好对应于 id-cf 在此例的步骤（4.2.）中会执行的重写。因此，该算法输出了公式（55）中的正确识别表达式。

算法第 (4.) 步的关键思想，尽管要精确阐述需要一些篇幅，却很简单：检查在任意 R-片段中，F_l 是否有任何变量 X 直观地说，在任何地方都必然被设为同一个值 x。这可能是因为该变量从另一个具有机制 c_X 的 R- 片段中复制而来，也可能是因为输入其中的状态各不相同却彼此一致，或者是因为它始终以此为条件 x。无论通过何种方式使该 X 在任何地方的取值都一致，只要一个 R-片段中的所有变量 X 都满足这一点，那么该 R-片段就是可识别的。

定理90.（id-cf的可靠性）：算法 id-cf 对反事实的识别是可靠的；也就是说，只要它没有输出失败，按 P_* 表示的输出 C 就是正确的反事实。

证明（概要）：见附录 E。 □

推论 91. 给定一个顶点为 O、属于 C 的祖先有向混合图 G，反事实项 W^k 和 $P_*(O)$，那么，与查询 W^k 对应的反事实 C 可由 G 和 P_* 识别，当且仅当对于 D，经 simplify-cf 简化后的对应未归一化图满足：

(α) 任意 λ_X 都被吸收到 c_X 中了。

(β) 在任意 R-片段 F_l 中，且对于任意 $X \in O$，条件 (4.1.1.a) 或 (4.1.1.b) 以及 (4.1.2.a) 或 (4.1.2.b) 均成立；也就是说，直观而言，F_l 中 X 的所有出现都具有一致的值。

现在显而易见的问题当然是算法 id-cf 的完备性。要证明完备性，必须证明：只要算法输出“失败”，也就是说，只要推论 91 的条件 (α) 和 (β) 至少有一个不成立，该反事实就不可识别；换言之，需要构造对 simplify-cf 输出的图 D 的两个不同解释，使它们都与 P_* 相容，却推出不同的反事实。因此，显然这需要比 C 是具有效应条件化的 cd-范畴更多的结构——我们需要“足够多的映射”，才能保证总能构造出反例。然而，在范畴 $\mathbf{Mat}_{\mathbb{R}^+}$ 中，完备性显然成立。首先，上述算法及其可识别性判据（推论 91）只是将参考文献 [41], 中的结果转化为我们的图示方法；该文献证明了可识别性与推论 91 中条件的等价性 —— 当然是用他们自己的术语表述的。其次，直观而言，在 $\mathsf{C} = \mathbf{Mat}_{\mathbb{R}^+}$ 的情形下，我们的表述也显然适用。上述两个不可识别性例子分别体现了这种直觉，涵盖了推论 91 的条件 (α) 或 (β) 不成立的两类情况。因此，为建立完备性所需的论证，本质上并非图示性质的。明确给出范畴 C 必须满足的精确条件，使文献中用于证明完备性的方法（例如参考文献 [41] 中的方法）能够适用，并且在我们的框架中变得显而易见，这或许并不困难，但我们将其留待未来研究。

注释92。（区分条件（α）和（β）的理由）：请注意，一方面，将给定祖先有向混合图中的双向边所表示的潜在共同原因变量，以及相应根化中的根节点 R，与另一方面，函数因果模型的背景噪声变量U_X 区别对待——后者每个变量仅有一个子节点 X ——是我们的选择，这不同于文献[41]中的典型表述。这一选择出于概念上的原因，并不意味着一般性有所不同。不难看出[37]，如果推论 91 的条件（α）不满足，即 simplify-cf 的输出仍明确包含背景噪声变量的状态，那么必然是因为对于某个 $X \in O$，有若干 f_X 以不同状态输入到 X 的某些父节点中（ λ_X除外）。因此，只需消除两类未观测变量之间的区别，即 R 中的变量与（U_X）中的变量$_{X \in R \cup O}$，推论 91 就会相对于整个未观测变量集合归结为条件（β）。本文倾向于保留这一区别，因为“世界内部”的未观测共同原因与“跨世界”的未观测共同原因，可以被视为导致可识别性失败的不同原因。始终会存在未观测的噪声变量，它们引入一定的随机性，并使得学习底层函数 f_X 变得不现实。因此，如果对于某个给定的因果结构，某个反事实由于条件（α）失效而不可识别，那么这可以与（β）失效区分开来——如果实际世界中的所有因果相关变量（即共同原因）都已被观测到，后者或许可以避免，而前者则在原则上更不现实。

备注 93。（祖先有向混合图和根化在不可识别性研究中的作用）：最后，回想一下备注 81 中提出的问题：在研究反事实的可识别性时，使用潜在投影 ADMG 是否会导致一般性的损失？据我们所知，文献中尚未回答这个问题。只要假设 C 具有适当的结构，从而能够证明推论 91 中条件的充分性，那么鉴于脚注 32，通过考察使用根化方法 ρ 而非 $\tilde{\rho}$对不可识别性造成的影响，这个问题或许很容易得到回答。

[37]通过反证法，并遵循 `simplify-cf` 中的拓扑序 π。

8.3 反事实的推广

最后，我们来讨论反事实概念的若干自然推广，这一概念见定义 78。这些推广分别扩展了反事实的两个主要特征：定义平行世界的干预，以及世界内部可能进行的条件化。

一般干预。首先，从平行世界模型推导反事实时，严格来说，没有理由坚持干预 σ_j 必须是 do-干预。正如上文所述，定义 78 中的 do-干预并无操作层面的意义，而只是形式化了为使某个变量取不同值所需的「小小奇迹」。因此，我们可以考虑通过任意平行世界模型定义的反事实；只要确定性开放模型 $\mathbb{F}$ 在干预 σ_j 下仍保持确定性，这些反事实便允许任意干预，而不局限于 do-干预。

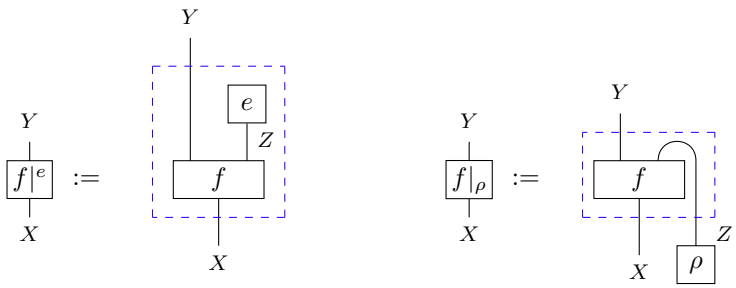
更进一步，我们可以考虑一种平行世界模型的泛化，其定义方式与（49）相同，但现在允许任意干预 $\sigma_{j\circ}$ 由此得到的模型 $\sigma_j(\mathbb{F})$ 可能不再是确定性的，因此每个世界的闭合模型 $\sigma_j(\mathbb{F}) \circ \mathbb{L}$ 也可能不再是函数因果模型。

对于这两种广义反事实，每个分支都可以看作一个假设中的不同世界：它与实际世界的差异在于某些变量的机制事实，而不只是这些变量所取值的事实。这种广义概念也可以对应于这样一种设想：大致来说，若某人实际并未进行干预 σ，却观测到了这样那样的情况，或许还观测到了此前进行的另一种不同干预 σ' 所产生的结果，那么如果当时进行了该干预，会发生什么。请注意，这类一般干预并未抹去「其他一切保持不变」这一点，因为构建广义平行世界模型时，仍然是从函数因果模型出发，并以相同方式共享 U 变量。

软条件化。反事实的第二个关键特征，是条件化发生在世界内部。到目前为止，我们只考虑了基于尖锐效应 $c^{(j)}$ 在 $C^{(j)}$ 上定义的、类似事件的“事实”进行条件化。然而，概率分布 P 不仅可以基于某个尖锐态 $X = x$ 进行条件化，从而得到$P(Y|X = x)$，也可以基于“一般态”（在 $\mathsf{Mat}_{\mathbb{R}^+}$ 中即分布）或 X上的效应（在 $\mathsf{Mat}_{\mathbb{R}^+}$ 中即正函数）所表示的“模糊事实”进行条件化。

因此，另一种自然的泛化是，考虑基于 $C^{(j)}$上任意状态或效应 $c^{(j)}$ 所给出的“软证据”进行条件化。这种模糊条件化相当于根据某些可能并不完备的知识，更新人们对各个世界中所取得数值的信念。

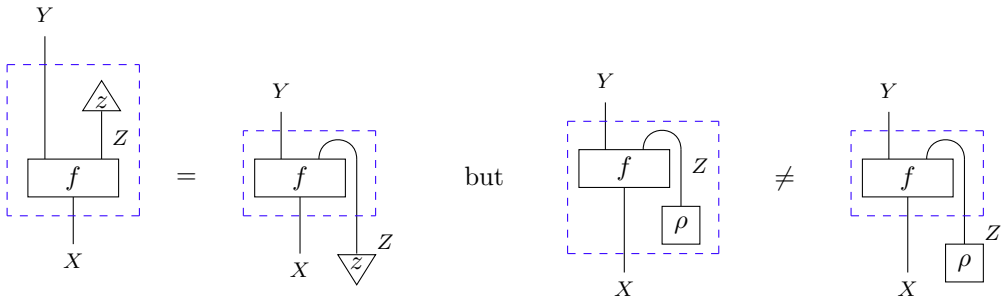
然而，概率论中一个未受足够重视的事实是，图示有助于阐明，实际上定义这种“软条件化” [100]有两种自然方式。给定一个态射 $f\colon X \to Y \otimes Z$，我们可以分别如下定义：基于 Z上的任意效应 e得到 f的上条件态 $f|^e$，以及基于 Z的任意状态 ρ 得到 f的下条件态 $f|_\rho$[38]。



在第2.5节中我们看到，当 z 是一个尖锐态且 $z^\dagger$ 是相应的尖锐效应时，两种更新过程一致，并由贝叶斯条件分布 $f|_z = f|^{z^\dagger}$给出。但对于一般态 ρ，令效应 e 由将 ρ 与一个帽复合得到，我们通常有 $f|^e \neq f|_\rho$，这是由于归一化框位置不同所致

[38]在没有可消去帽的范畴中，$f|_\rho$ 中的虚线方框应替换为更一般的条件信道概念，具体讨论见附录C。

归一化框的。



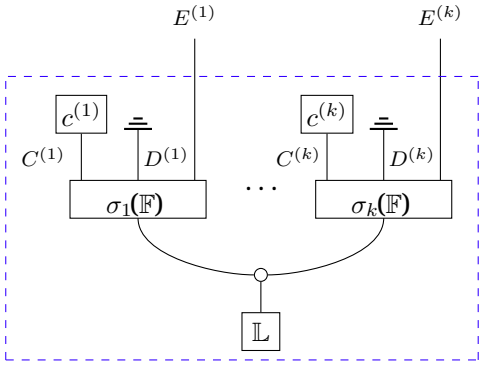
Jacobs 对软条件化或信念更新的两种形式都进行了区分和深入探讨（其中 f 是一个状态，且 e 位于图示之外） [100, 101], ，并分别称之为“Pearl 式”和“Jeffrey 式”更新。

例 94。在 $\mathbf{Mat}_{\mathbb{R}^+}^{+}$ 中，两种更新形式由以下公式给出。

$$f|^{e}(y \mid x) := \frac{\sum_{z \in Z} e(z) f(y, z \mid x)}{\sum_{y' \in Y, z' \in Z} e(z') f(y', z' \mid x)} \qquad f|_{\rho}(y \mid x) := \sum_{z \in Z} \frac{\rho(z) f(y, z \mid x)}{\sum_{y' \in Y} f(y', z \mid x)}$$

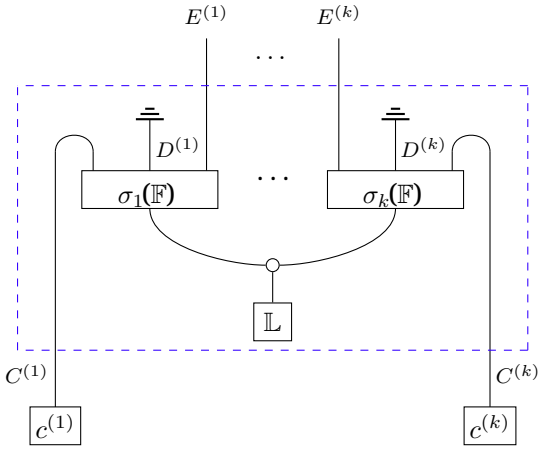
现在我们来讨论相应的两种广义反事实形式，它们允许进行一般干预，并可通过一般效应或状态进行条件化。

定义 95。设 $\mathbb{M} = \mathbb{F} \circ \mathbb{L}$ 是 $\mathbf{C}$ 中的一个函数因果模型，其确定性部分为 $\mathbb{F}$，输出为 O；设 σ_1，...，σ_k是施加于 $\mathbb{F}$的任意干预，并将所得模型的输出标记并划分为 $O^{(j)} = C^{(j)} \cup D^{(j)} \cup E^{(j)}$，使得至少对于某个 $j \neq j'$，有 $C^{(j)} \neq \emptyset \neq E^{(j')}$。由效应 $c^{(1)}$，...，$c^{(k)}$ 确定的广义反事实是如下形式的状态：



(61)

来自状态 $c^{(1)}, \ldots, c^{(k)}$ 的广义反事实是一种如下形式的状态：



(62)

关于此类广义反事实，还有许多值得进一步探讨的内容，包括对每个世界上的一般干预 σ_j 的解释（以及是否可能将其排除），以及选择哪种形式的条件化（61）、（62）具有最自然的反事实解释。原则上，在同一个反事实状态中混合不同类型的条件化也并非不可设想。我们也将更广义的反事实概念的可识别性判据留待未来研究。

9 展望

我们在一般的范畴论框架内，以弦图语言介绍了因果模型框架中的许多关键概念和工具。其中包括因果模型、函数因果模型、干预和反事实，以及不那么常见的开放因果模型概念。这拓展了 Fong [7] 和 Jacobs 等人 [9, 11] 的工作，使我们能够以图示方式处理因果效应和反事实的可识别性问题。主要结论之一是，图示推理是进行因果推理最自然、最简便的方式——因果推理似乎天生就适合用弦图术语来表述。

未来的研究方向有很多（尽管本文篇幅已经很长）。下面列出其中一些，并非面面俱到。

因果效应的识别。一个重要目标是推导出因果效应可识别性的一般图示解法。首先，这需要找到一个图示条件，使其等价于如下形式的因果效应的可识别性： $P(Y|Z;\mathrm{do}(X))$，其中 X、Y、Z 是变量的任意不相交子集；此外，还需要通过图示推理推导出如何识别 $P(Y|Z;\mathrm{do}(X))$。这将推广文献 [11], 中讨论的 JKZ 结果（见第7节），并与传统文献中针对该问题已有的可靠且完备的算法相对应（例如，参见文献 [41]）。

正如第7.3节所述，并在脚注 22 中有所暗示，这里一种可能的做法是，在对图进行“手术”后，寻找一种图示条件，类似于 Fritz 等人提出的 d-分离的范畴泛化。事实上，一种选择是从运用他们的结果入手，以便以图示方式表达并推导 do-演算规则。

其次，人们希望针对非原子干预 σ_X，推导出 $P(Y|Z;\sigma_X)$ 可识别性的一般解，因为这类干预在图解处理中既十分自然，也具有很强的实际意义。本文仅讨论了这类一般干预的具体示例。一种直接的选择是将参考文献 [37] 中的 σ-演算规则表示为图示形式。

因果模型的高阶映射视角。为了研究因果效应可识别性，因果模型的分裂节点表示很可能是一个特别合适的视角，附录 F 对此作了概述。该表示不采用解释网络图，而是使用梳形映射，也称为断路，或高阶映射。Jacobs 等人 [11] 的研究以及参考文献 [102], 和 [74], 中都已出现梳形映射；后者还首次提出了“分裂节点因果模型”这一术语。其基本思路是“切断”因果模型网络图中的导线，从而打开一个“槽位”，以便在其中插入干预。尽管它们在数学上是不同的对象，但彼此密切相关——因果模型的分裂节点表示与其网络图表示一一对应，而且在两者之间转换十分容易。除附录 F 外，本文没有详细介绍分裂节点形式体系，因为本文关注的是仅凭网络图便可轻松处理的方面。不过，我们并不认为本文的阐述与以分裂节点为核心的阐述有本质区别。相反，这两种视角相辅相成，分别适用于不同目的。因此，值得进一步明确：除了模型的定义之外，因果模型框架的哪些方面可以通过分裂节点因果模型得到自然体现，以及这两种视角各自的优势究竟在哪里。

连续条件化与模糊条件化。正如附录C中所指出的，我们希望将定义条件信道 $P(Y \mid X)$ 的方法——以及借此拓展第7节和第8节中的应用——推广为一种更一般的范畴化条件化形式，以适用于附录A所概述的连续情形和测度论情形。为此，我们可以尝试为态射的条件分布指定图形公理

$f|_Z$： $X \otimes Z \to Y$，从而扩展本文为归一化给出的公理，并解决文献 [10] 中提出的一个开放问题。

我们还可以用另一种方式推广这些条件分布，第8.3节中提到过这一点：允许基于“模糊”事实进行条件化，这些事实由一般态和效应给出，而非仅由像$Z = z$这样的尖锐事实给出。鉴于集合 Z 在 $P(Y \mid Z; \text{do}(X))$ 中是否为空会影响该因果效应的可识别性，人们自然会想知道，如果允许基于此类模糊事实进行条件化，已知的可识别性条件会有何变化。

反事实。正如第8.2.3节所指出的，显而易见的下一步是证明反事实识别算法 id-cf 的完备性。这将确立推论 91 中图示可识别性判据的必要性，并完成对 Shpitser 等人发表于文献 [41]中的结果的图解翻译。确立完备性之后，很自然地会进一步考察：弦图对反事实的处理，是否确实能让我们轻松论证，基于祖先有向混合图（ADMG）作为潜在投影来研究反事实的可识别性并不失一般性。这将推广 Richardson 等人发表于文献 [38] 中的一项关于因果效应的类似结果（参见备注81和93），并填补传统文献中目前似乎存在的一处空白。

除此之外，值得研究我们的方法是否可以自然地扩展，以涵盖所谓的嵌套反事实 [103, 80, 104] ，方法是允许通过共享的背景变量 U以外的方式在不同世界之间建立联系。最后，第8.3节中对反事实概念的泛化涉及以模糊事实进行条件化，以及由比 do-干预更一般的干预所定义的平行世界，这值得进一步探索。特别是，可以研究在这种情况下，反事实的识别标准和方法会如何变化。

因果模型的概念。本文仅讨论了无环因果结构。然而，文献中已有对循环因果结构的研究，尤其是为了描述反馈回路（例如，参见参考文献 [105, 106, 107, 108]）。一个自然的下一步是探索此类循环因果模型的弦图表示。进一步而言，将它们与参考文献 [75], 中研究的（经典情形下的）循环因果模型联系起来，也会很有启发性；后者原则上允许存在性质截然不同的循环性——并非源自可在时间序列中展开的反馈回路——且与量子理论基础中的“不定因果序”研究相关。

另一个值得探索的方向是如何处理潜变量模型。本文引入并使用了祖先有向混合图的根化（见第 6 节），即使用一种有向无环图，使其对应的潜在投影还原出给定的祖先有向混合图。在许多情况下，研究可识别性问题时，这并不会造成任何一般性的损失。不过，人们不禁会想，是否存在一种更直接的图示表示——无论是字符串图表示还是其他形式——能够基于祖先有向混合图构建模型，并使图中表达的结构性质等价于相对于该祖先有向混合图的嵌套马尔可夫条件，或者更一般地，基于某个混合有向无环图 [91, 90]构建模型。

开放因果模型的结构。在第5节中，我们引入了 cd-范畴 C 中开放因果模型的范畴 OCM(C)，但指出它的结构不止是一个幺半范畴，例如，它还能够“深入”一个态射，将其内部变量外化。若能准确确定开放因果模型本身的形式结构，以及“开放因果结构”（即网络图，或等价地，开放有向无环图）的形式结构，将会很有意义。后者与开放图 [39, 40] 以及 Fritz 和 Liang 意义上的“广义因果模型” [35]的形式结构有关。

变换与自动化因果推理。我们介绍了若干干预以及对（开放）因果模型更一般的变换，但还有许多其他变换有待探索。更一般地说，我们希望建立一种形式化方法，用于推理模型的有效变换，同时注意到每种变换都可能要求输入模型满足某些条件（例如，确保结果仍然无环）。一种方法是定义开放模型变换所构成的范畴，其中对象是满足特定标准的模型集合，而态射则是各种变换。我们希望这有助于自动因果推理和开放模型的构建：通过形式化我们如何将同一变换（例如 $\text{do}(X = x)$）应用于一系列模型（例如，任何包含X 作为变量的模型），或者将粗粒化和因果抽象的概念形式化为变换。

如引言中所述，这项工作的长期目标之一是助力开发此类方法，以自动学习和发现数据的因果表征，正如所谓的“因果表征学习” [19, 46, 45]。

基础层面。如引言中所述，当前工作有望有助于理解因果性的若干其他基础层面。这包括将因果模型置于参考文献 [71] 中提出的因果推断理论框架内。这还包括探索量子因果模型的一种潜在图示表示 [72, 74, 77, 76]，并据此探讨“量子反事实” [109] 。

10 致谢

我们谨此感谢 Bob Coecke、Steve Clark、Matty Hoban、KonstantinosMeichanetzidis、Ilya Shpitser、Robin Evans 和 Ciar´an Gilligan-Lee 提供的宝贵讨论和反馈。

参考文献

[1] H. Beebee、C. Hitchcock 和 P. Menzies，《牛津因果关系手册》。牛津大学出版社，2009年。[2] J. D. Gallow，“因果关系的形而上学”，载于《斯坦福哲学百科全书》，E. N. Zalta 和 U. Nodelman 编。形而上学研究实验室，斯坦福大学，2022年秋季版，2022年。[3] C. Hitchcock，“概率性因果关系”，载于《斯坦福哲学百科全书》，E. N. Zalta 编。形而上学研究实验室，斯坦福大学，2021年春季版，2021年。[4] J. Pearl，《因果性》。剑桥大学出版社，2009年。[5] P. Spirtes、C. Glymour 和 R. Scheines，《因果关系、预测与搜索》。麻省理工学院出版社，第2版，2000年。[6] B. Coecke 和 R. W. Spekkens，“经典与量子贝叶斯推断的图示”，Synthese 186卷第3期（2012年），651–696页。[7] B. Fong，“因果理论：贝叶斯网络的范畴论视角”，arXiv预印本 arXiv:1301.6201（2013年）。[8] K. Cho 和 B. Jacobs，“通过弦图实现分解与贝叶斯反演”，《计算机科学中的数学结构》29卷第7期（2019年），938–971页。[9] B. Jacobs、A. Kissinger 和 F. Zanasi，“通过弦图手术进行因果推断”，载于软件科学与计算结构基础国际会议，313–329页，施普林格，2019年。[10] T. Fritz，“马尔可夫核、条件独立性及充分统计量定理的综合方法”，《数学进展》370卷（2020年），107239。[11] B. Jacobs、A. Kissinger 和 F. Zanasi，“通过弦图手术进行因果推断：干预与反事实的图示方法”，《计算机科学中的数学结构》31卷第5期（2021年），553–574页。[12] T. Fritz 和 A. Klingler，“范畴概率中的 d-分离准则”，arXiv预印本 arXiv:2207.05740（2022年）。[13] T. Fritz 和 A. Klingler，“范畴概率中的 d-分离准则”，《机器学习研究杂志》24卷第46期（2023年），1–49页。[14] E. Bareinboim、J. D. Correa、D. Ibeling 和 T. F. Icard，“论珀尔层级与因果推断基础”，2021年。

[15] B. Glocker、M. Musolesi、J. Richens 和 C. Uhler，“数字医学中的因果关系”，《自然·通讯》12卷第1期（2021年）。[16] H. Reynaud、A. Vlontzos、M. Dombrowski、C. Lee、A. Beqiri、P. Leeson 和 B. Kainz，“达达尼昂：反事实视频生成”，arXiv预印本 arXiv:2206.01651（2022年）。[17] A. Vlontzos、B. Kainz 和 C. Lee，“通过深度孪生网络估计范畴反事实”，

.

[18] C. M. Gilligan-Lee、C. Hart、J. Richens 和 S. Johri，《利用有向因果发现检测因果对中的潜在共同原因》，《IEEE神经网络与学习系统汇刊》（2022年）1–0。[19] B. Sch¨olkopf、F. Locatello、S. Bauer、N. R. Ke、N. Kalchbrenner、A. Goyal 和 Y. Bengio，《迈向因果表征学习》，《IEEE会议录》109卷第5期（2021年）612–634。[20] E. Wolfe、R. W. Spekkens 和 T. Fritz，《利用潜变量进行因果推断的膨胀技术》，《因果推断期刊》7卷第2期（2019年）。[21] E.-C. Boghiu、E. Wolfe 和 A. Pozas-Kerstjens，《Inflation：用于经典与量子因果相容性的Python库》，arXiv预印本 arXiv:2211.04483（2022年）。[22] S. Mac Lane，《工作数学家的范畴论》，第5卷。施普林格科学与商业媒体，2013年。[23] J. Baez 和 M. Stay，《物理学、拓扑学、逻辑与计算：罗塞塔石碑》。施普林格，2011年。[24] P. Selinger，《幺半范畴图形语言综述》，《物理学的新结构》（2011年）289–355。[25] S. Abramsky 和 B. Coecke，《量子协议的范畴语义》，载于《第19届IEEE计算机科学逻辑年度研讨会论文集》，2004年，第415–425页，IEEE。2004年。[26] B. Coecke 和 A. Kissinger，《量子过程图解：量子理论与图示推理入门》，载于《图示表示与推理：第10届国际会议，Diagrams 2018，爱丁堡，英国，2018年6月18日至22日，论文集》第10卷，第28–31页，施普林格。2018年。[27] B. Coecke 和 R. Duncan，《相互作用的量子可观测量：范畴代数与图示方法》，《新物理学杂志》13卷第4期（2011年）043016。[28] B. Fong 和 D. I. Spivak，《应用范畴论导论：组合性的七个草图

y.

剑桥大学出版社，2019年。[29] D. Shiebler、B. Gavranovi´c 和 P. Wilson，“机器学习中的范畴论”，arXiv 预印本 arXiv:2106.07032（2021年）。[30] N. Ghani、J. Hedges、V. Winschel 和 P. Zahn，“组合博弈论”，载于第33届年度 ACM/IEEE计算机科学逻辑研讨会论文集，第472–481页。2018年。[31] B. Coecke、M. Sadrzadeh 和 S. Clark，“组合分布式意义模型的数学基础”，《语言分析》36（2010年）345–384页。[32] B. Jacobs 和 F. Zanasi，“贝叶斯学习的谓词/状态变换器语义”，《理论计算机科学电子札记》325（2016年）185–200页。[33] B. Jacobs 和 F. Zanasi，“贝叶斯推理中影响的形式语义”。[34] B. Jacobs 和 F. Zanasi，“贝叶斯推理的逻辑要义”，《概率编程基础》（2019年）295–331页。[35] T. Fritz 和 W. Liang，“自由 gs-幺半范畴与自由马尔可夫范畴”，arXiv 预印本 arXiv:2204.02284（2022年）。

[36] B. Jacobs、A. Kissinger 和 F. Zanasi，“通过弦图手术进行因果推断：一种图示 ic
干预与反事实的方法”，《计算机科学中的数学结构》31卷第5期，（2021年）553–574页。[37] J. Correa和E. Bareinboim，“随机干预的演算：因果效应识别 a
nd
替代实验”，载于《AAAI人工智能会议论文集》，第34卷，页码10093–10100。2020年。[38] T. S. Richardson、R. J. Evans、J. M. Robins和I. Shpitser，“无环有向混合图的嵌套马尔可夫性质”，2017年。[39] B. Fong，“装饰余跨”，arXiv预印本 arXiv:1502.00872（2015年）。[40] J. C. Baez和K. Courser，“结构化余跨”，arXiv预印本 arXiv:1911.04630（2019年）。[41] I. Shpitser和J. Pearl，“因果层级的完全识别方法”，《机器学习研究杂志》9（2008年）1941–1979页。[42] J. Pearl和D. Mackenzie，《为什么之书》。Basic Books，纽约，2018年。[43] B. Coecke和S. Gogioso，《图解量子》。Quantinuum，2022年。[44] J. Pearl，“因果革命的七点启示：机器学习的理论障碍”，arXiv预印本 arXiv:1801.04016（2018年）。[45] B. Sch¨olkopf和J. von K¨ugelgen，“从统计学习到因果学习”，arXiv预印本 arXiv:2204.00607（2022年）。[46] B. Sch¨olkopf，“机器学习中的因果性”，arXiv预印本 arXiv:1911.10500（2019年）。[47] G. Parascandolo、N. Kilbertus、M. Rojas-Carulla和B. Sch¨olkopf，“学习独立因果机制”，载于国际机器学习会议论文集，页码4036–4044，PMLR。2018年。[48] Y. Bengio、T. Deleu、N. Rahaman、R. Ke、S. Lachapelle、O. Bilaniuk、A. Goyal和C. Pal，“用于学习解缠因果机制的元迁移目标”，arXiv预印本 arXiv:1901.10912（2019年）。[49] I. Dasgupta、J. Wang、S. Chiappa、J. Mitrovic、P. Ortega、D. Raposo、E. Hughes、P. Battaglia、M. Botvinick和Z. Kurth-Nelson，“基于元强化学习的因果推理”，arXiv预印本 arXiv:1901.08162（2019年）。[50] K. Chalupka、P. Perona和F. Eberhardt，“视觉因果特征学习”，arXiv预印本 arXiv:1412.2309（2014年）。[51] K. Chalupka、F. Eberhardt和P. Perona，“多层级因果效应系统”，载于人工智能与统计学会议论文集，页码361–369，PMLR。2016年。[52] K. Chalupka、T. Bischoff、P. Perona和F. Eberhardt，“利用微观层面气候数据进行因果特征学习，以无监督方式发现厄尔尼诺现象”，arXiv预印本 arXiv:1605.09370（2016年）。[53] F. Eberhardt，“绿色与格鲁因果变量”，Synthese 193卷第4期，（2016年）1029–1046页。[54] K. Chalupka、F. Eberhardt和P. Perona，“因果特征学习：综述”，Behaviormetrik
a
44卷第1期（2017年），137–164页。[55] E. Bengio、V. Thomas、J. Pineau、D. Precup 和 Y. Bengio，“独立可控特征”，arXiv 预印本 arXiv:1703.07718（2017年）。[56] F. Locatello、S. Bauer、M. Lucic、G. Raetsch、S. Gelly、B. Sch¨olkopf 和 O. Bachem，“挑战无监督学习解缠表示中的常见假设”，载于《国际机器学习会议》，第4114–4124页，PMLR，2019年。

[57] F. Leeb、G. Lanzillotta、Y. Annadani、M. Besserve、S. Bauer 和 B. Sch¨olkopf，“通过架构实现结构：无需正则化的解缠表示”，arXiv 预印本 arXiv:2006.07796（2020年）。[58] F. Locatello、B. Poole、G. R¨atsch、B. Sch¨olkopf、O. Bachem 和 M. Tschannen，“毫不妥协的弱监督解耦”，载于《国际机器学习会议》，第6348–6359页，PMLR，2020年。[59] J. Mitrovic、B. McWilliams、J. Walker、L. Buesing 和 C. Blundell，“通过不变因果机制进行表示学习”，arXiv 预印本 arXiv:2010.07922（2020年）。[60] N. R. Ke、J. Wang、J. Mitrovic、M. Szummer、D. J. Rezende 等，“神经因果表示的摊销学习”，arXiv 预印本 arXiv:2008.09301（2020年）。[61] X. Shen、F. Liu、H. Dong、Q. Lian、Z. Chen 和 T. Zhang，“解耦生成式因果表示学习”，arXiv 预印本 arXiv:2010.02637（2020年）。[62] Y. Wang 和 M. I. Jordan，“表示学习的期望：因果视角”，arXiv 预印本 arXiv:2109.03795（2021年）。[63] F. Tr¨auble、E. Creager、N. Kilbertus、F. Locatello、A. Dittadi、A. Goyal、B. Sch¨olkopf 和 S. Bauer，“从相关数据中学得的解缠表示”，载于《国际机器学习会议》，第10401–10412页，PMLR，2021年。[64] P. Lippe、S. Magliacane、S. L¨owe、Y. M. Asano、T. Cohen 和 S. Gavves，“CITRIS：从时间干预序列中实现因果可识别性”，载于《国际机器学习会议》，第13557–13603页，PMLR，2022年。[65] J. Brehmer、P. De Haan、P. Lippe 和 T. Cohen，“弱监督因果表示学习”，arXiv 预印本 arXiv:2203.16437（2022年）。[66] D. Shiebler、B. Gavranovi´c 和 P. Wilson，“机器学习中的范畴论”，arXiv 预印本 arXiv:2106.07032（2021年）。[67] “人工智能范畴（系列讲座）。” https://cats.for.ai/.[68] T. Cohen，“迈向具身人工智能的扎根因果理论”，arXiv 预印本 arXiv:2206.13973（2022年）。[69] E. F. Rischel 和 S. Weichwald，“组合抽象误差与因果模型范畴”，载于《人工智能中的不确定性》，第1013–1023页，PMLR，2021年。[70] D. Janzing 和 S. H. G. Mejia，“现象学因果性”，arXiv 预印本 arXiv:2211.09024（2022年）。[71] D. Schmid、J. H. Selby 和 R. W. Spekkens，“理清因果与推断的纠葛：因果推断理论框架”，arXiv 预印本 arXiv:2009.03297（2020年）。[72] J.-M. A. Allen、J. Barrett、D. C. Horsman、C. M. Lee 和 R. W. Spekkens，“量子共同原因与量子因果模型”，《物理评论X》7卷（2017年7月），031021。 https://link.aps.org/doi/10.1103/PhysRevX.7.031021。[73] F. Costa 和 S. Shrapnel，“量子因果建模”，《新物理学杂志》18卷第6期（2016年），063032。[74] J. Barrett、R. Lorenz 和 O. Oreshkov，“量子因果模型”，arXiv 预印本 arXiv:1906.10726（2019年）。[75] J. Barrett、R. Lorenz 和 O. Oreshkov，“循环量子因果模型”，《自然·通讯》1卷第1期（2021年），1–15页。

2

[76] N. Ormrod、A. Vanrietvelde 和 J. Barrett，《扇区约束存在时的因果结构
，
及其在量子开关中的应用》，arXiv 预印本 arXiv:2204.10273（2022年）。[77] R. Lorenz 和 J. Barrett，《酉变换的因果与复合结构》，Quantum 5（2021年）511。[78] A. Vanrietvelde、H. Kristj´ansson 和 J. Barrett，《路由量子电路》，Quantum 5（2021年）503。[79] J. Tian 和 J. Pearl，《因果效应的一般识别条件》。eScholarship，加利福尼亚大学，2002年。[80] J. Pearl，《反事实的算法化》，《数学与人工智能年鉴》61卷第1期（2011年），29–39页。[81] T. Fritz，《关于马尔可夫核、条件独立性及充分统计量定理的综合方法》，《数学进展》370（2020年）107239。[82] B. Coecke，《向执业物理学家介绍范畴》，载于《什么是范畴论》，第45–7页

4,
Polimetrica Monza。2006年。[83] C. Hitchcock 和 M. R´edei，《赖欣巴赫共同原因原理》，载于《斯坦福哲学百科全书》，E. N. Zalta 编。形而上学研究实验室，斯坦福大学，2020年春季版，2020年
.

[84] R. Lorenz，《量子因果模型：赖欣巴赫原理的精神对于理解量子因果结构的价值》，Synthese 200卷第5期，（2022年）424。[85] J. Otsuka 和 H. Saigo，《因果过程理论：概述》，2022年10月。http://philsci-archive.pitt.edu/21267/。[86] J. Peters、D. Janzing 和 B. Sch¨olkopf，《因果推断要素：基础与学习算法》。麻省理工学院出版社，2017年。[87] T. VERMA 和 J. PEARL，《因果网络：语义与表达能力》，载于《人工智能中的不确定性》，R. D. SHACHTER、T. S. LEVITT、L. N. KANAL 和 J. F. LEMMER 编，第9卷《机器智能与模式识别》，第69–76页。North-Holland，1990年。http://www.sciencedirect.com/science/article/pii/B9780444886507500111。[88] D. Geiger、T. Verma 和 J. Pearl，《识别贝叶斯网络中的独立性》，Networks 20卷第5期，507–534，https://onlinelibrary.wiley.com/doi/pdf/10.1002/net.3230200504。https://onlinelibrary.wiley.com/doi/abs/10.1002/net.3230200504。[89] D. Janzing 和 S. H. G. Mejia，《现象学因果性》，arXiv预印本 arXiv:2211.09024（2022年）。[90] R. J. Evans，《离散贝叶斯网络的边缘分布》，《统计年鉴》46卷第6A期，（2018年）2623–2656。[91] R. J. Evans，《贝叶斯网络边缘分布图》，《斯堪的纳维亚统计学杂志》43卷第3期，（2016年）625–648。[92] R. M. Byrne，《理性想象：人们如何创造现实的替代方案》。麻省理工学院出版社，2007年。[93] W. Starr，《反事实》，载于《斯坦福哲学百科全书》，E. N. Zalta 和 U. Nodelman 编。形而上学研究实验室，斯坦福大学，2022年冬季版，2022年。[94] A. Balke 和 J. Pearl，《反事实概率：计算方法、界限与应用》，载于《1994年不确定性会议论文集》，第46–54页。爱思唯尔，1994年。[95] J. Y. Halpern 和 J. Pearl，《原因与解释：结构模型方法。第一部分：原因，

”
《英国科学哲学杂志》56卷第4期（2005），843–887。

[96] J . Y. Halpern and J. Pearl, “Causes and explanations: A structural-model approach. part ii : 解释》，《英国科学哲学杂志》56卷第4期，（2005）889–911。[97] C. Avin、I. Shpitser 和 J. Pearl，《路径特定效应的可识别性》。[98] D. Lewis，《反事实》。美国马萨诸塞州剑桥：布莱克威尔，1973年。[99] D. Lewis，《因果关系》，《哲学杂志》70卷第17期，（1973）556–567。[100] B. Jacobs，《通过杰弗里更新规则或珀尔更新规则改变信念的数学》，《人工智能研究杂志》65（2019年）783–806。[101] B. Jacobs，《从正确之处学习与从错误之处学习》，arXiv预印本 arXiv:2112.14045（2021年）。[102] I. Friend 和 A. Kissinger，《量子过程中因果影响的识别》。量子物理与逻辑会议论文集。[103] J. Correa、S. Lee 和 E. Bareinboim，《从任意替代实验中进行嵌套反事实识别》，载于《神经信息处理系统进展》，M. Ranzato、A. Beygelzimer、Y. Dauphin、P. Liang 和 J. W. Vaughan 编，第34卷，第6856–6867页。Curran Associates, Inc.，2021年。
https://proceedings.neurips.cc/paper/2021/file/36bedb6eb7152f39b16328448942822b-Paper

.pdf.

[104] J. Pearl，《直接效应与间接效应》，载于《第十七届人工智能不确定性会议论文集》，UAI’01，第411–420 页。Morgan Kaufmann 出版社，2001年。[105] P. Spirtes，《反馈模型的有向循环图表示》，载于《第十一届人工智能不确定性会议论文集》，UAI’95，第 491–498 页。Morgan Kaufmann 出版社，1995年。[106] T. Richardson，《有向循环图的马尔可夫等价刻画》，《国际近似推理》17 卷第 2-3 期，（1997年）107–162 页。[107] P. Forr´e 和 J. M. Mooij，《含循环和潜变量的图模型的马尔可夫性质》，arXiv 预印本 arXiv:1710.08775（2017年）。[108] P. Forr´e 和 J. M. Mooij，《存在循环、潜在混杂因素和选择偏差时的因果演算》，载于《人工智能中的不确定性》，第 71–80 页，PMLR。2020年。[109] A. K. Suresh、M. Frembs 和 E. G. Cavalcanti，《量子因果模型中反事实的语义》，arXiv 预印本 arXiv:2302.11783（2023年）。[110] P. Panangaden，《概率关系》，伯明翰大学计算机科学学院研究报告 CSR（1998年）59–74 页。[111] J. Pearl，《智能系统中的概率推理：合理推断网络》。Morgan Kaufmann，1988年。[112] S. Lauritzen，《有向马尔可夫性质》。牛津大学，讲义，2011年。http://www.stats.ox.ac.uk/~steffen/teaching/gm11/dag.pdf。[113] G. Chiribella、G. M. D’Ariano 和 P. Perinotti，《量子网络的理论框架》，Phys. Rev. A 80（2009年8月）022339。https://link.aps.org/doi/10.1103/PhysRevA.80.022339。[114] O. Oreshkov、F. Costa 和 ˇC. Brukner，《无因果序的量子关联》，《自然·通讯》3 卷第 1 期，（2012年）1092。[115] S. Uijlen 和 A. Kissinger，《因果结构的范畴语义》，《计算机科学中的逻辑方法》15（2019年）。[116] A. Bisio 和 P. Perinotti，《高阶量子理论的理论框架》，《皇家学会会刊 A》475 卷第 2225 期，（2019年）20180706。

[117] M. Wilson 和 G. Chiribella, 《高阶过程理论中的因果性》, arXiv 预印本 arXiv:2107.14581（2021年）。

[118] M. Wilson 和 G. Chiribella, 《物理过程变换的数学框架》, arXiv 预印本 arXiv:2204.04319（2022年）。

[119] M. Wilson、G. Chiribella 和 A. Kissinger, 《量子超映射由局域性刻画》, arXiv 预印本 arXiv:2205.09844（2022年）。

A 测度论概率

cd-范畴的框架使人们能够考虑比 $\mathbf{Mat}_{\mathbb{R}^+}$ 更为一般的情形。特别是，它们为一般可测空间之间的概率通道提供了图解处理，例如由连续概率密度函数定义的概率通道。在假设读者具备一些测度论知识的前提下，我们在此简要说明如何将测度论概率视为 cd-范畴或马尔可夫范畴。更多细节请参见例如参考文献 [110, 8, 10]。

以下示例范畴并非本文其余部分所必需，因为本文的示例聚焦于有限情形 $\mathbf{Mat}_{\mathbb{R}^+}$ 和 FStoch，这些范畴足以说明因果模型的关键概念。

例 96. *[110]* 在 cd-范畴 Prob 中，对象是可测空间 (X, Σ_X)，也就是带有 σ-代数 Σ_X 的集合 X，其元素称为 X 的可测子集。回忆一下，X 上的次概率测度是一个函数 $\mu\colon \Sigma_X \to [0,1]$，满足

$$\mu\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} \mu(E_k)$$

只要 $E_k \in \Sigma_X$ 两两不相交即可。态射 $f\colon X \to Y$ 由从 X 到Y的马尔可夫次核给出。它是一个函数，将每个 $x \in X$ 映射到 Y上的次概率测度 $f(x)$，并且对每个 $A \in \Sigma_Y$，映射 $x \mapsto f(x, A) := f(x)(A)$ 都是可测函数 $X \to [0,1]$。

恒等态射 id_X 将每个 x 映射到点测度 δ_x，其中 $\delta_x(A) = 1$ 当且仅当 $x \in A$，否则为 0。$f\colon X \to Y$ 与 $g\colon Y \to Z$ 的复合由积分给出：

$$(g \circ f)(x, A) := \int_{y \in Y} g(y, A)\, df(x)(y)$$

张量由 $X \otimes Y = X \times Y$ 定义，并配备乘积 σ-代数 $\Sigma_{X \otimes Y}$。对于子核$f\colon X \to Y$和 $g\colon W \to Z$，我们定义 $f \otimes g$，使其将每个 $(x, y) \in X \times Y$映射为乘积测度 $f(x) \otimes g(y)$。这里 $I = \{\star\}$ 是一个单元素集。

一个状态 $\omega\colon I \to X$ 可以通过 $\omega = \omega(\star)$ 识别为定义在 X 上的一个次概率测度 ω；类似地，一个效应 $e\colon X \to I$ 相当于一个可测函数 $e = e(\star)\colon X \to [0,1]$。特别地，取值于 $\{0,1\}$ 且非零的效应 e，可通过其示性函数 $e = 1_M$ 识别为可测子集 $M \subseteq X$，其中

$$1_M(x) = \begin{cases} 1 & x \in M \\ 0 & otherwise \end{cases}$$

丢弃效应 $\top_X = 1_X$ 对应于子集 X，其定义为 $x \mapsto 1$ 对所有 $x \in X$。标量是一个概率 $p \in [0,1]$。

这里，状态 ω 是规范化的，当且仅当它是一个概率测度，且 $\omega(X) = 1$。更一般地，子核$f\colon X \to Y$是信道，当且仅当它是概率通道，即将每个 $x \in X$ 映射为概率测度 $f(x)$。信道的马尔可夫子范畴通常记作 Stoch。

状态（或过程）的边缘化对应于概率论通常意义上的边缘化。也就是说，对于 $f\colon X \to Y \times Z$，边缘分布 $g\colon X \to Y$由 $g(x, A) = f(x, A \times Z)$ 给出。

每个点 $x \in X$ 都通过概率测度 δ_x 在 X 上定义一个状态，其中当且仅当 $x \in A$ 时，$\delta_x(A) = 1$，否则 $\delta_x(A) = 0$。复制映射对每个 $x \in X$定义为 $x \mapsto \delta_{(x,x)}$。因此，每个状态 δ_x 都是尖锐的，而任意（部分）函数 $f\colon X \to Y$ 都可通过以下方式视为确定性（部分）信道 $f\colon X \to Y$：当 $f(x)$ 有定义时，$f(x) = \delta_{f(x)}$；否则，$f(x) = 0$。

Prob 具有归一化，对于态射 $f\colon X \to Y$，其定义为

$$\mathrm{norm}(f)(x,A) := \begin{cases} \frac{f(x,A)}{f(x,Y)} & f(x,Y) \neq 0 \\ 0 & otherwise \end{cases}$$

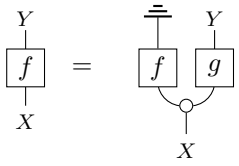
不过，Prob 一般不具有帽。要得到帽，可以在子 cd-范畴 BorelProb 中进行，其对象是标准 Borel 空间。对于任意这样的空间 X，对每个 $x \in X$，子集 $\{x\}$ 都是可测的，因此通过示性函数 $x^\dagger = 1_{\{x\}}$ 确定了 X 上的一个尖锐效应。此外，对角子集 $\Delta = \{(x,x) \mid x \in X\}$ 是可测的，从而得到帽 $= 1_\Delta$。不过请注意，尽管这里存在帽，但它们一般并不可消去。特别地，对于 X 上的一个状态 ω，将 ω 与帽复合所得到的 X 上的效应为 $e(x) := \omega(\{x\})$，它可以为零而 $\omega = 0$ 不成立。

B 归一化

B.1 归一化的唯一性

根据定义 8，cd-范畴 C 中的归一化只要存在，实际上就是唯一的。为此，我们引入以下概念。

定义 97. 对于任意态射 $f\colon X \to Y$，若态射 $g\colon X \to Y$ 满足以下条件，我们就称它将 f 归一化：



(63)

因此，部分信道是一个将自身归一化的态射。

引理 98. 归一化具有以下性质：

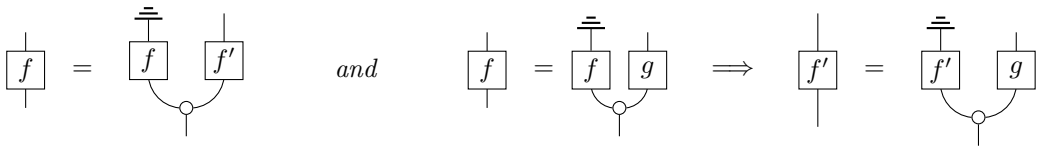
1. 传递性：若 g 将 f 归一化，且 h 将 g 归一化，则 h 将 f归一化。
2. 反对称性：若 f 和 g 互相归一化，则 $f = g$。

证明。 记 e_f 为通过复制出一条额外导线并将其与 $\circ f$复合得到的态射。那么 g将 f归一化，当且仅当 $f = g \circ e_f$。然后通过应用丢弃可知， $e_f = e_g \circ e_f$也成立。对于 (1)，注意到$f = g \circ e_f = h \circ e_g \circ e_f = h \circ e_f$。对于 (2)，注意若 f和 g互相归一化，则 f将自身归一化。现在 $f = g \circ e_f = f \circ e_g \circ e_f = f \circ e_f \circ e_g = f \circ e_g = g$，利用复制的交换性来交换 e_f和 e_g。

□

一般而言，使态射 f归一化的态射 g有很多个，因为直观地说，当输入 x 满足 $f(x) = 0$时， g 可以在输入上取任意值。 f的“正确”归一化也应将这类输入映射为 0。刻画这种“极小性”的另一种方式如下。

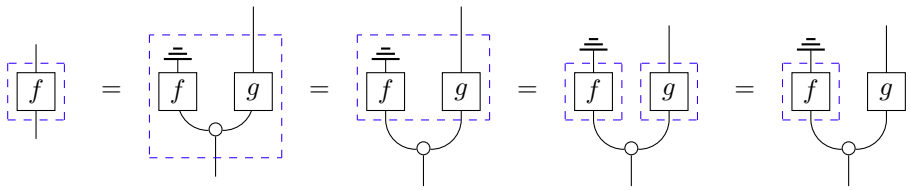
定义 99. 如果存在一个将 f 归一化的部分信道 f'，并且对于任何其他将 f 归一化的部分信道 g， g也将 f' 归一化，则称态射 f 具有最小归一化。



由引理 98 可知，最小归一化一旦存在就是唯一的。根据定义，如果 f 本身已经是一个部分信道，那么它满足必要条件，因此等于其最小归一化。

命题 100. 设 C 是一个带有归一化的 cd-范畴。那么 norm(f) 是 f 的最小归一化。

证明. 根据 (10)，norm(f) 确实对 f 进行了归一化；我们必须验证它是最小的。但是，如果 g 是一个对 f 进行归一化的部分信道，那么 g 会对 norm(f) 进行归一化，因为：



□

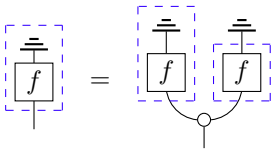
由极小规范化的唯一性，我们得到如下推论。

推论 101. cd-范畴 C 上的归一化结构若存在，则唯一。

鉴于命题 100，自然会问，是否存在更多初等条件，能够确保极小规范化满足定义 8 的性质。

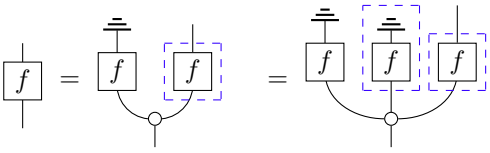
引理 102. 假设 C 中每个态射都有一个极小规范化。那么极小规范化满足 (12)。

证明。 对于任意态射 f，容易看出，由于 norm(f) 是一个部分信道，因此 $\circ\,\mathrm{norm}(f)$ 也是如此。对 (63) 两边应用丢弃可得 $\circ\,\mathrm{norm}(f)$ 归一化了 $\circ\, f$。因此，它也归一化了 norm($\circ\, f$)，即：

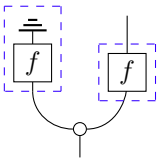


(64)

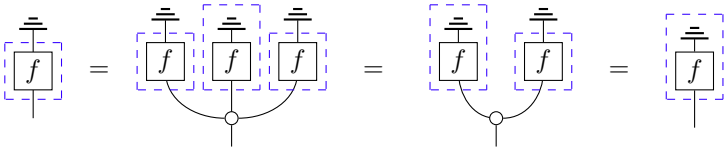
现在，对于所有态射 f，我们有



因此， f 由态射归一化



可以验证，该态射是一个部分信道。更一般地，若 g、h 满足条件，则 $(g \otimes h)$ 也满足条件。因此，这个态射也必须对 norm(f) 进行归一化。将其写成等式并应用丢弃，可得：

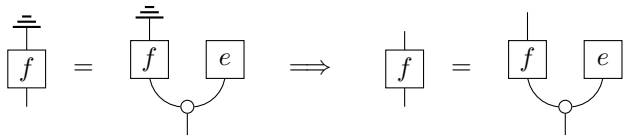


最后一步使用了 (64)。 □

推论 103. cd-范畴 C 具有归一化，当且仅当每个态射都有一个最小归一化，且这些最小归一化共同满足 (11) 和 (13)。

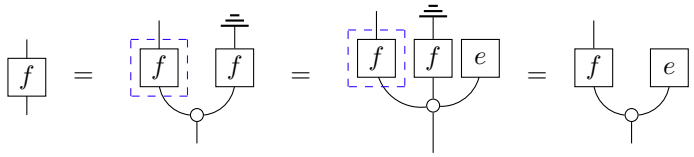
注意，归一化的存在还确保满足以下条件。

引理 104：如果 C 具有归一化，那么 C 还满足引理 9 的性质以及丢弃消去律：



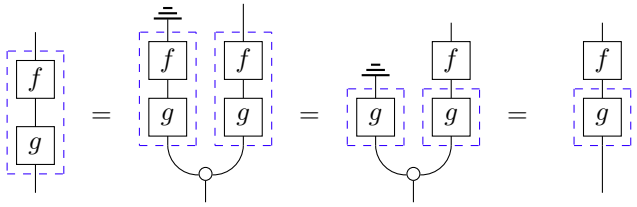
对所有态射 f 和效应 e。

证明。我们先考虑丢弃消去。假设左侧的等式成立。 n

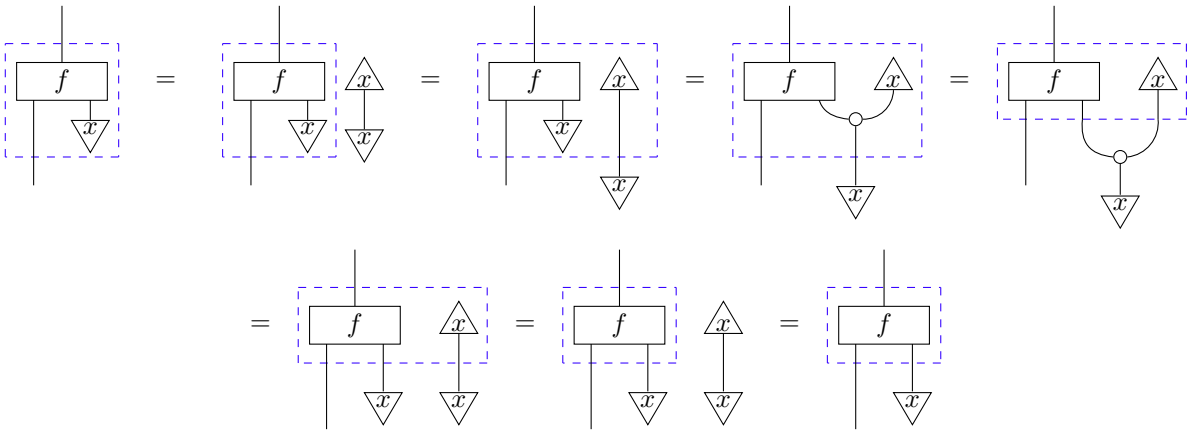


接下来，我们证明引理 9 的条件 (4)、(5)、(6)。

对于 (4)，容易看出，当 f 是一个信道时，$f \circ$范数(g) 是一个部分信道，并且将 $f \circ g$ 归一化。因此，它也必然将范数$(f \circ g)$ 归一化，因为后者是 $f \circ g$ 的最小归一化。但于是



对于 (5)，注意到是一个~~信~~道，因此它是一个部分信道，所以范数$(\ ) = $。将~~=~~$f$ 替~~换~~为 $f \otimes$并利用 (11) 即得~~=~~所需结果。然后如下面所示可推出 (6)，这里用到每个效应 $x^\dagger$ 都是确定性的，因此也是一个部分信道。



□

B.2 推导归一化

还可以通过其他一些自然的方式推导出归一化的存在。

或许，理解归一化最直接的方式是考察它对确定性态的作用，如式 (9) 所示。回忆一下，当对所有对象 X、Y，都存在一个态射 $0 : X \to Y$，使得对于所有态射 f，$0 \circ f = 0 = f \circ 0$ 和 $f \otimes 0 = 0$ 时，我们称 C 具有零态射。

如果对于所有确定性状态 $f \circ x = g \circ x\, x$，我们都有 $f = g$，那么我们称一个 cd-范畴具有足够多的确定性状态。如果对于任意状态 ω、ω'和非零标量 r，只要 $r \cdot \omega = r \cdot \omega'$，就有 $\omega = \omega'$，那么我们称 C 具有标量消去性。最后，如果每个非零状态 ω 都存在一个归一化态 norm(ω)，使得等式 (7) 成立（由标量消去性可知该态是唯一的），并以虚线方框表示，那么我们称 C 具有态归一化。

命题 105. 设 C 是一个具有标量消去性、态归一化和足够多确定性状态的 cd-范畴。那么，C 具有归一化，当且仅当对于每个态射 f，都存在一个态射 $\mathrm{norm}(f)$，使得对于所有确定性状态 x，式 (9) 成立。

证明。根据假设，范数 $\mathrm{norm}(f)$ 由以下事实唯一确定：对于所有确定性状态 x，$\mathrm{norm}(f)\circ x = \mathrm{norm}(f\circ x)$；而根据引理 9 (6)，这一条件是必要的。

现在根据态归一化的定义，(10) 对所有输入确定性态 x 都成立，因此根据假设也成立。由态归一化的唯一性可知，只要 ω 已归一化，就有 $\mathrm{norm}(\omega)=\omega$；由此可得，对于所有 f，$\mathrm{norm}(f)=\mathrm{norm}(\mathrm{norm}(f))$，因此 $\mathrm{norm}(f)$ 始终是一个部分信道。

反过来，如果 f 是一个部分信道，那么将确定性状态 x 代入部分信道的定义，我们可知 $f\circ x=(\circ f\circ x)(f\circ x)$。根据标量消去性，要么 $f\circ x=0$，要么 $\circ f\circ x=1$，即 $f\circ x$ 已归一化。因此，对于所有确定性状态 x，都有 $f\circ x=\mathrm{norm}(f\circ x)=\mathrm{norm}(f)\circ x$，从而得到 $f=\mathrm{norm}(f)$。

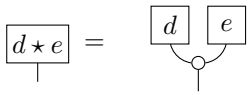
现在我们来验证 (11)、(12)、(13)。

(11)：容易证明，在任意 cd-范畴中，$X\otimes Y$ 的每个确定性状态都可写成 $x\otimes y$，其中 X、Y 的确定性状态 x、y 分别如此。将这种形式的状态代入等式两边，可知该条件化简为：对于所有状态 a、b，范数$(a\otimes b)$ $=$ 范数 (a) $\otimes$ 范数 (b)；这可由归一化的唯一性推出。

(12)：代入确定性状态后，要求对于所有状态 ω，范数$(\ \circ\omega)=\circ$范数(ω) 这是因为根据态归一化，范数$(\ \circ\omega)=1$ 或为 0；在后一种情况下，$\circ$范数$(\omega)=1$ 也为 0。

(13)：代入确定性状态 x，我们有范数$(f)\circ\circ\circ x=$范数$(f)\circ(x\otimes x)=$范数$(f\circ(x\otimes x))=$范数$(f\circ\curlyvee\circ x)=$范数$(f\circ)\circ x$。 □

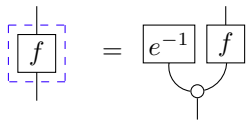
另一种推导归一化的方法，是利用对标量和效应进行“求逆”的能力，具体如下。对于效应 d,e，我们将效应 $d\star e$ 记作



那么，效应 e 是一个 部分信道，当且仅当它是确定性的，当且仅当 $e=e\star e$；而当且仅当 $d=d\star e$ 时，效应 d 会归一化。

定义 106. 设 C 为一个 cd-范畴。若对于每个效应 e，都存在唯一的效应 e^{-1}，使得 $e\star e^{-1}=\mathrm{norm}(e)=\mathrm{norm}(e^{-1})$ 是 e 和 e^{-1} 的最小归一化，则称 C 具有效应逆。

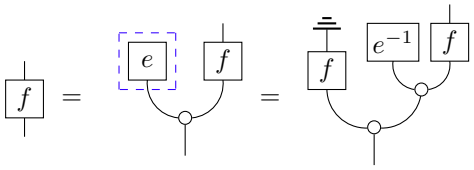
命题107. 设 C 为具有效应逆且满足丢弃消去的 cd-范畴。那么 C 具有归一化；对于态射 f，其归一化由下式给出



(65)

其中 $e=\circ f$。

证明。我们首先证明 (65) 是 f 的最小归一化。很容易验证它是一个部分信道。现在，如果 g 归一化 f: $X\to Y$，那么将 (63) 的两边与态射 $X\to X$（通过复制出 X 并与 e^{-1} 复合而得）复合，可得 g 也归一化 (65)。反过来，我们必须证明 (65) 归一化 f。对范数(e) 归一化 e 这一事实应用丢弃消去，可知

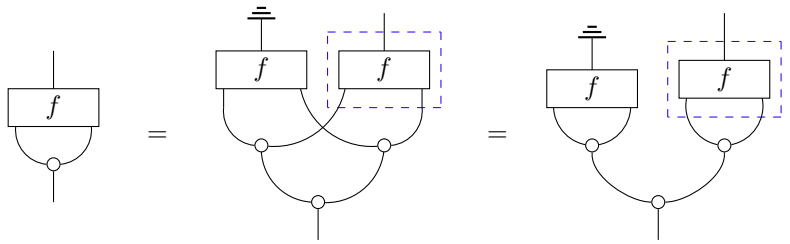


因此，(65) 确实构成了 f 的最小归一化。根据推论 103，只需验证 (11) 和 (13)。

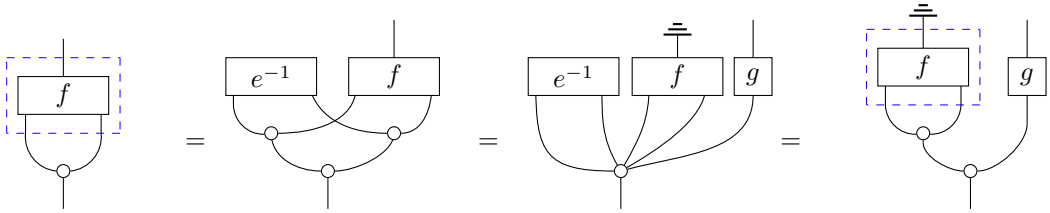
现在验证 (11)。由 (65) 可知，这等价于验证对于任意效应 d、e，$(d \otimes e)^{-1} = d^{-1} \otimes e^{-1}$。由于 $(d^{-1} \otimes e^{-1}) \star (d \otimes e) =$ 范数$(d) \otimes$ 范数(e)，根据逆元的唯一性，只需证明范数$(d \otimes e) =$ 范数$(d) \otimes$ 范数(e)，也就是说，只需在 $f = d$ 和 $g = e$ 是效应的情况下证明 (11)。根据定义，范数$(d) \otimes$ 范数(e) 将 $d \otimes e$ 归一化。反过来，如果某个效应 k 将 $d \otimes e$ 归一化，那么我们有

$$\text{范数}(d) \otimes \text{norm}(e) = (d^{-1} \otimes e^{-1}) \star (d \otimes e) = (d^{-1} \otimes e^{-1}) \star (d \otimes e) \star k = \text{norm}(d) \otimes \text{norm}(e) \star k$$

and so 范数$(d) \otimes$ 范数(e) 构成最小归一化，正如所要求的。因此，(11) 成立。最后，我们必须验证 (13)。首先，注意到范数 $(f) \circ$ 确实归一化了 $f \circ$，因为：



假设 g 是另一个将 $f \circ$ 归一化的部分信道。那么我们有



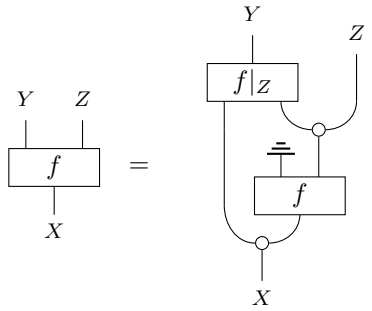
正如所要求的。 □

例108。$\text{Mat}_{\mathbb{R}^+}$ 满足命题 105 和命题107 的所有要求，因此既具有归一化，也具有效应的逆元。

C 条件化与条件独立性

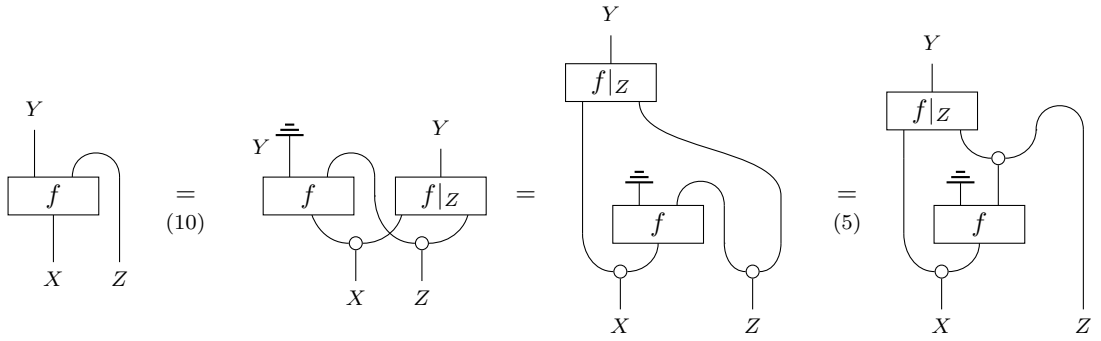
让我们更准确地看看，我们对条件化的处理与文献中先前的方法有何关系。下面的内容推广了引理 12。

命题109. 设 C 是一个具有效应条件化的 cd-范畴。那么每个条件分布 $f|_Z$ 都满足以下性质。



(66)

证明. 利用归一化的性质，我们得到以下结论。



应用帽子的可消去性即可得出结论。 □

Cho 和 Jacobs 将归一化联合态 ω 定义为：当存在信道$\omega|_X$ 使得式 (17) 成立时， ω 具有分解。更一般地，Fritz 指出，若对每个态射 $f\colon\ X \to Y \otimes Z$，都存在一个满足式 (66) 的态射 $f|_Z$，则马尔可夫范畴 C 具有条件分布。 [10]请注意，要使该 $f|_Z$作为信道唯一存在， f 在 X上必须具有全支撑。 Z

因此，当存在可消去帽时，我们的条件化概念将条件化推广到了全支撑情形之外，即使 $f|_Z$ 只是部分信道（而非要求 f 为信道），也能唯一确定它。然而，当帽不可消去时，这一概念可能不同于标准的概率条件化概念，如下所示。

例 1 10。在 BorelProb 中，考虑定义在 X、Y 上的联合概率测度 ω ，其中 $X = \mathbb{R}^n$、 $Y = \mathbb{R}^m$，由密度 $p(x, y)$ 确定，使得

$$\omega(A \times B) = \int_{x\in A}\int_{y\in B} p(x,y)dxdy$$

对于 $B \in \Sigma_Y$，我们定义

$$\omega(x, B) := \int_{y\in B} p(x,y)dy$$

Then ω 确实具有一个条件分布 $\omega|_X$ ，其含义如式 (66) 所示（即给定 b 的一个分解 17 y

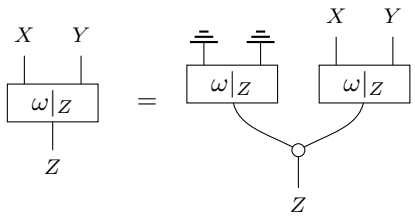
$$\omega|_X(x, B) := \frac{\omega(x, B)}{\omega(x, Y)}$$

当分母非零时如此，否则为零。然而，将 ω 与一个帽复合得到的效应 e 为 $e(x) := \omega(\{x\} \times B) = 0$，因为单元素集 $\{x\}$ 的勒贝格测度为 0。因此，按照式 (15) 定义的条件分布$\omega|'_X$ 就是零态。

因此，只有当帽是可消去的时，式 (15) 才适合作为条件信道的定义；而在连续情形中，这一条件不成立。若能扩展本文的方法，例如在更一般的 cd-范畴中处理条件信道 $P(Y|X)$，将会很有意思。为此，既希望将条件分布从信道推广到部分信道，也希望采用一种类似于我们处理归一化时所用方法的图形化处理。

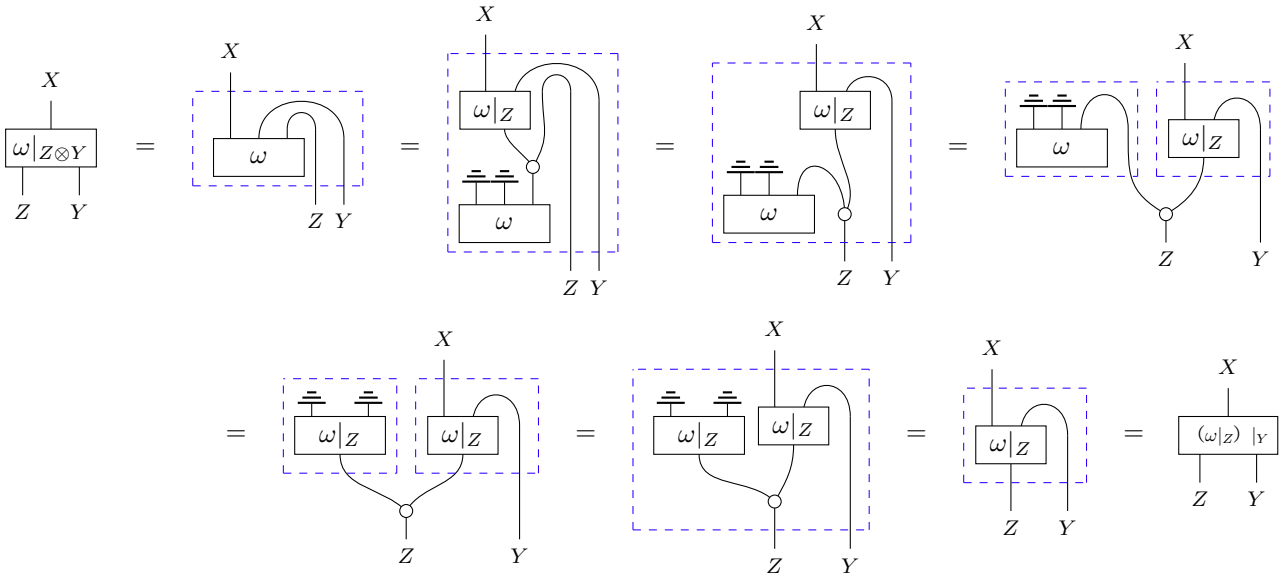
C.1 命题 15 的证明

证明。设 C 是一个具有效应条件化的 cd-范畴，且 ω 是 $X \otimes Y \otimes Z$ 的一个状态。首先注意，由于 $\omega|_Z$ 是一个部分信道，因此可得



(67)

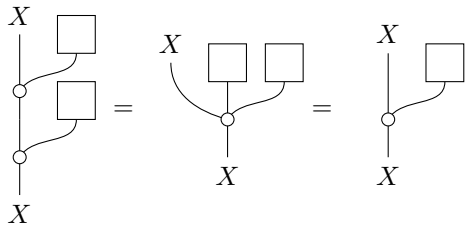
然后，利用引理 12、方程（5）和定义 8



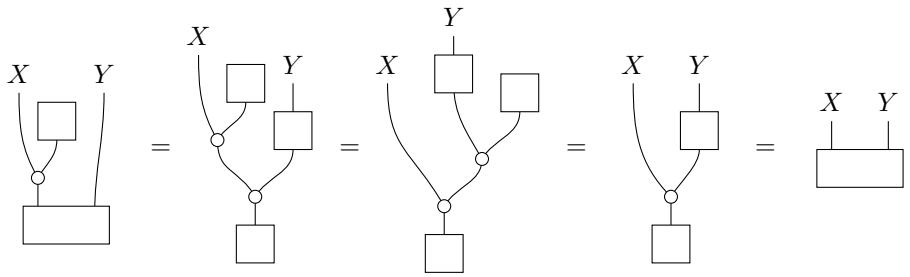
□

C.2 引理 16 的证明

证明。对于（1），没有 注意，每个支撑效应都是部分信道，因此是确定性的。所以我们有 e



方程（2）也可由以下事实推出： $\omega|_X$ 是一个部分信道。于是（3）成立，因为:

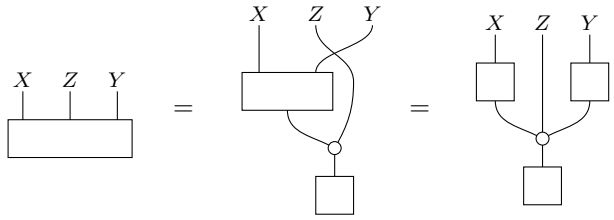


□

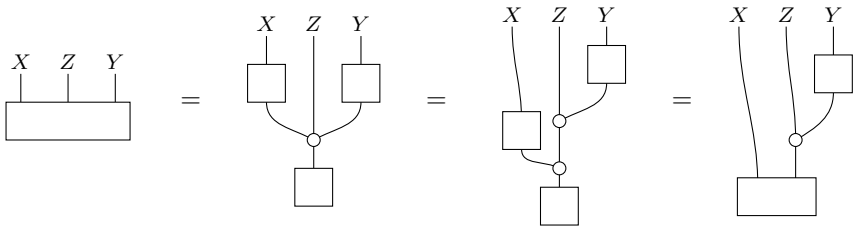
C.3 引理 18 的证明

证明。设 C 是一个具有效应条件化的 cd-范畴，且 ω 是 $X \otimes Y \otimes Z$的一个状态。

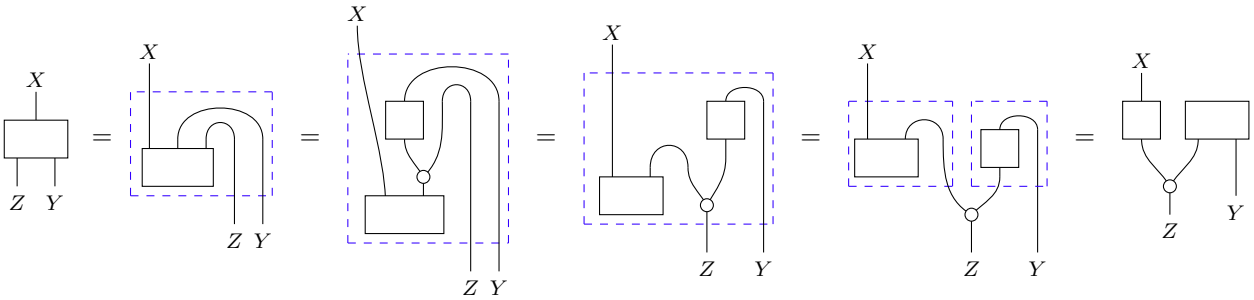
“$(X \perp\!\!\!\perp Y|Z)_\omega \Rightarrow$ (1)”:



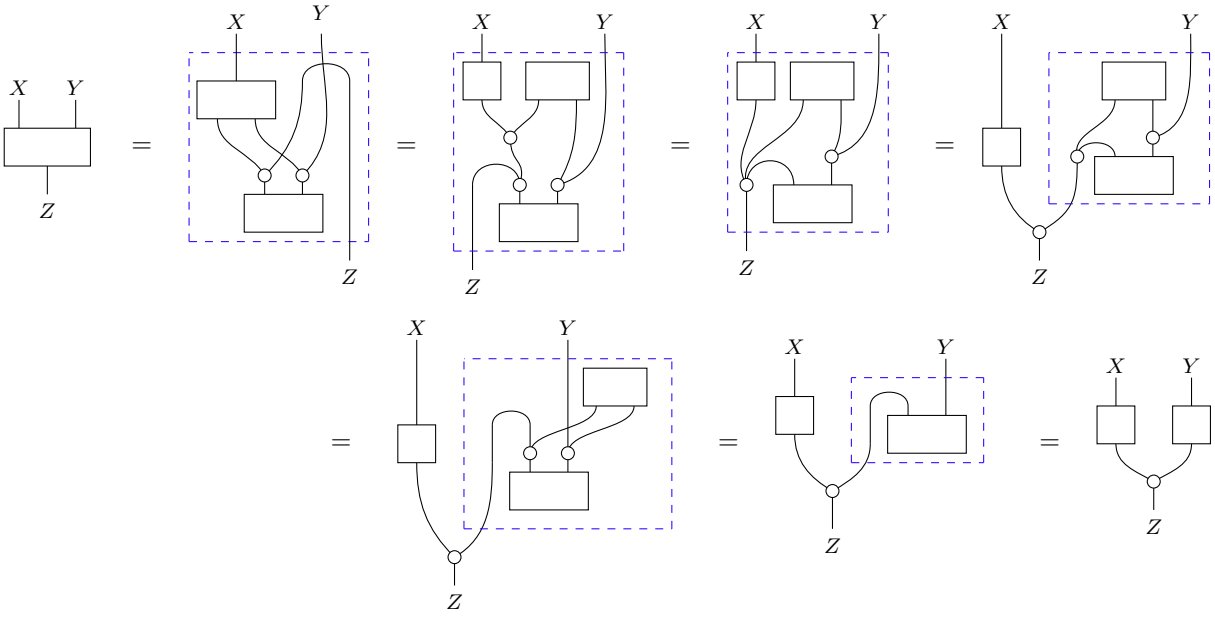
“(1) $\Rightarrow$ (2)”:



“(2) $\Rightarrow$ (3)”：利用 (2)、等式 (5) 和定义 8，可得



“(3) $\Rightarrow (X \perp\!\!\!\perp Y|Z)_\omega$”：利用引理 12、(3)、等式 (5) 和定义 8，可得



□

C.4 定理 19 的证明

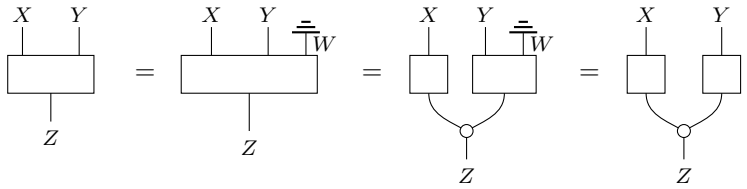
证明。设 C 为一个带有 效应条件化的 cd-范畴，且 ω 是 $X \otimes Y \otimes W \otimes Z$ 的一个状态。半拟阵公理如下：

$$
\begin{array}{rll}
\text{Symmetry:} & (X \perp\!\!\!\perp Y|Z)_\omega & \Leftrightarrow \quad (Y \perp\!\!\!\perp X|Z)_\omega \\
\text{Decomposition:} & (X \perp\!\!\!\perp Y \otimes W|Z)_\omega & \Rightarrow \quad (X \perp\!\!\!\perp Y|Z)_\omega \ \wedge \ (X \perp\!\!\!\perp W|Z)_\omega \\
\text{Weak Union:} & (X \perp\!\!\!\perp Y \otimes W|Z)_\omega & \Rightarrow \quad (X \perp\!\!\!\perp Y|Z \otimes W)_\omega \\
\text{Contraction:} & (X \perp\!\!\!\perp Y|Z \otimes W)_\omega \ \wedge \ (X \perp\!\!\!\perp W|Z)_\omega & \Rightarrow \quad (X \perp\!\!\!\perp Y \otimes W|Z)_\omega
\end{array}
$$

现在，我们通过直接计算逐一验证这些公理。

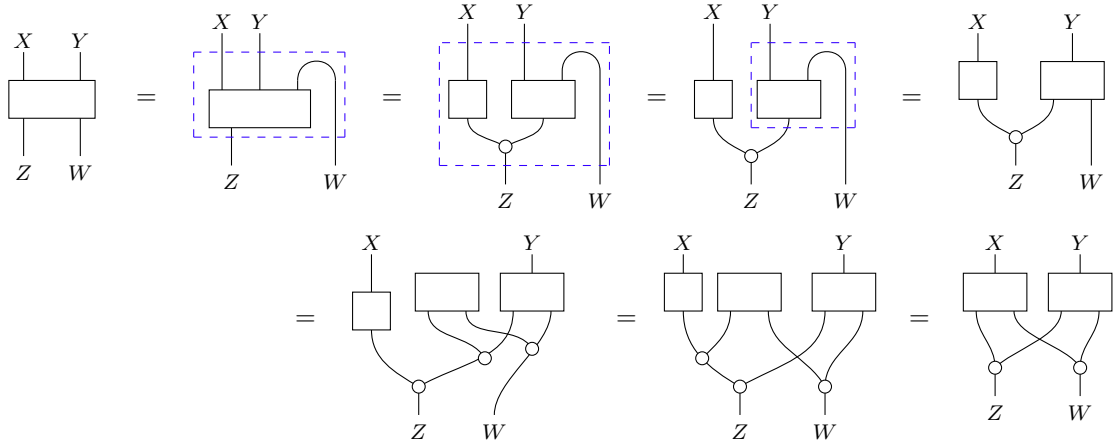
对称性：显然。

分解：假设 $(X \perp\!\!\!\perp Y \otimes W|Z)_\omega$，则



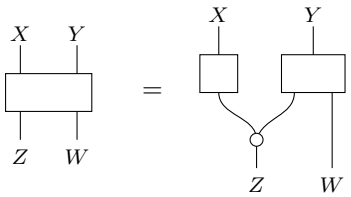
因此，$(X \perp\!\!\!\perp Y|Z)_\omega$。$(X \perp\!\!\!\perp W|Z)_\omega$ 的情形类似。

弱并：假设 $(X \perp\!\!\!\perp Y \otimes W|Z)_\omega$。那么，利用定义 8、引理 16，以及 $(X \perp\!\!\!\perp W|Z)_\omega$（根据分解公理，该式由假设推出），可得

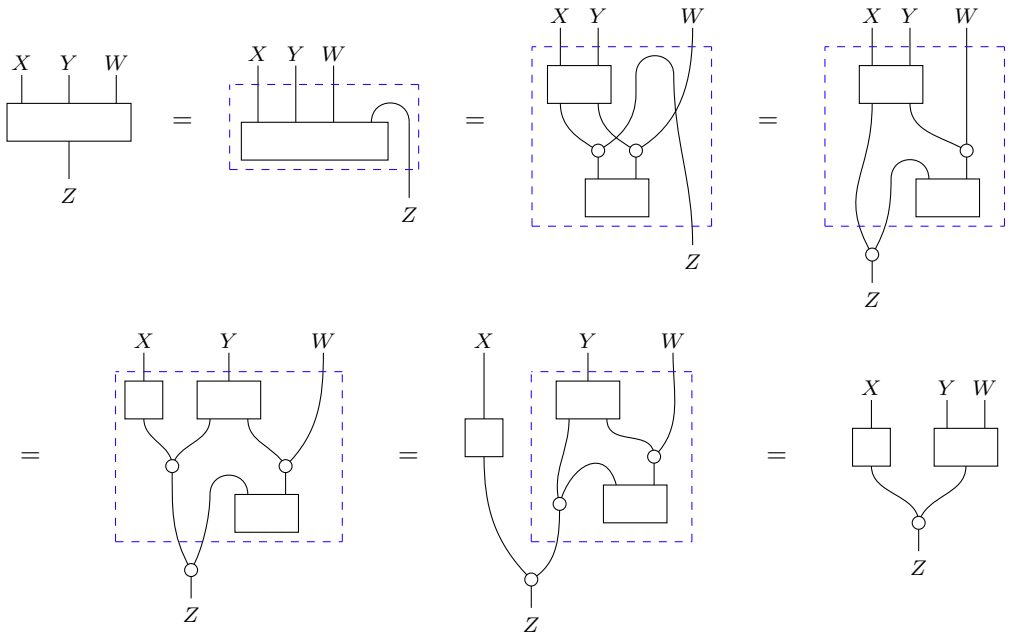


因此，$(X \perp\!\!\!\perp Y|Z \otimes W)_\omega$。

收缩：假设 $(X \perp\!\!\!\perp Y|Z \otimes W)_\omega$ 且 $(X \perp\!\!\!\perp W|Z)_\omega$。结合弱并公理，并按照上述证明该公理时所采用的类似步骤，可以直接得到



然后计算



因此，$(X \perp\!\!\!\perp Y \otimes W | Z)_{\omega}$。 □

C.5 定理 40 的证明

为确保完备性并便于查阅，我们先给出 d-分离的定义。

定义 111。（阻断路径与 d-分离 [111]）：给定一个有向无环图 G，节点 N 和 N'之间的一条路径，如果满足以下任一条件，则被节点集合 W 阻断：

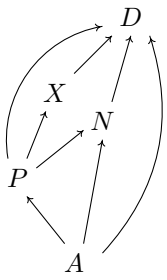
1. *a chain* $A \to M \to B$ *or a fork* $A \leftarrow M \to B$ *with the middle node* $M \in W$
2. 存在一个碰撞点 $A \to C \leftarrow B$，使得 C 及其所有后代均不属于 W。

对于节点 Y、Z和 W的子集，若对于每个 $N \in Y$和 $N' \in Z$，N与 N'之间的每条路径都被 W阻断，则称 Y 和 Z被 W d-分离，记作 $(Y \perp\!\!\!\perp Z | W)_G$。

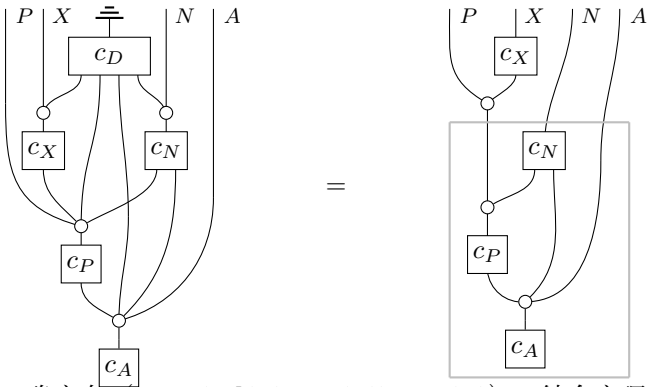
请注意，定理 19 已证明，三元关系 $(\perp\!\!\!\perp \_ | \_)_{\omega}$ 满足半拟阵公理。众所周知，这意味着局部马尔可夫条件与全局马尔可夫条件等价；后者不过是 d-分离对相应三元关系 [87, 112] 的可靠性。因此，剩下要做的就是验证：给定一个顶点为 V 的有向无环图 G，任何在 V 上、按照网络图 $D_{G,\ O=V}$ 进行因子分解的状态 ω，都满足局部马尔可夫条件，即 $\forall X \in V$ 有$\big(X \perp\!\!\!\perp (\mathrm{Nd}(X) \setminus \mathrm{Pa}(X)) \mid \mathrm{Pa}(X)\big)_{\omega}$，其中 $\mathrm{Nd}(X)$ 表示 X 的非后代。

定理 40 的证明。给定一个有向无环图 G，其顶点为 V，假设 ω 是 V 上的一个状态，并且它根据网络图 $D_{G,\ O=V}$ 进行因子分解。令 $X \in V$，写作 $P := Pa(X)$，并定义 $D := V \setminus \{X\} \setminus Nd(X)$（即后代），再将 $Nd(X) \setminus Pa(X) =$划分为：$A \cup N$，其中 A 是 X 的祖先（不包括父节点），而 N 是其余的非后代。如果将顶点集合 D、P、A 内部的结构“压缩”，并且

N，有向无环图 G 可视为以下图的子图：



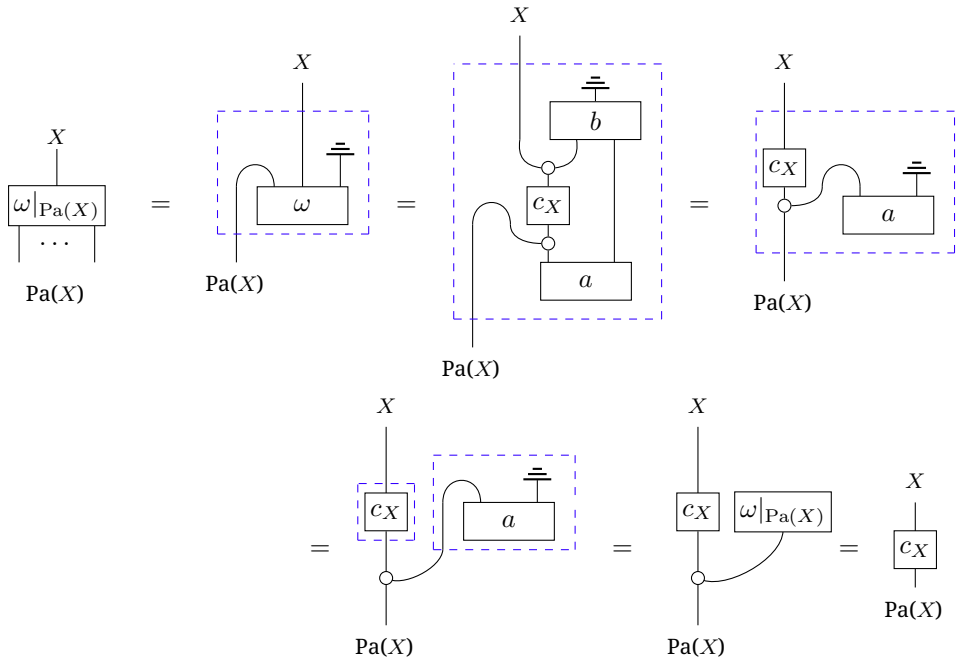
鉴于 ω 等于对带信道的 $D_{G,O=V}$ 的一种解释，丢弃 D 后 ω 的边缘态可写为：



因此，根据引理 18 的 (2)，确实有 $\left(X \perp\!\!\!\perp (\mathrm{Nd}(X) \setminus \mathrm{Pa}(X)) \mid \mathrm{Pa}(X)\right)_{\omega}$。结合定理 19，即可得证。□

C.6 引理 42 的证明

由于它在 $D_{\mathbb{M}}$ 上的因子分解，存在信道 a、b，使得 ω 如下进行因子分解。

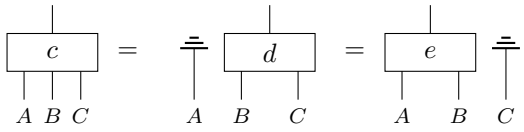


D 修剪干预

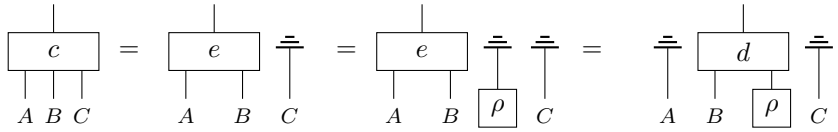
以下结果使我们能够从信道中移除不必要的输入，使其成为忠实信道，从而定义修剪干预。

引理112. 假设 c 是一个具有输入 S 的信道，且每个输入对象至少有一个归一化态。那么存在唯一的子集 $T \subseteq S$ 和信道 $d = \text{trim}(c)$，使得式（29）成立，且 d 是忠实的。

证明. 式（29）中的态射 d 自动是唯一的，因为它是将任意归一化态代入 c 中每个 T 输入所得。先考虑 c 有三个输入 A、B、C，且 c 的输出分别独立于 A 和 C 的情形。此时有



于是，对 C 的任意归一化态 ρ，我们得到



因此，输出独立于 $\{A, C\}$，或者等价地说，独立于 $A \otimes C$。将这一结论迭代应用于陈述中的一般信道 c，可知输出独立于 T 当且仅当输出独立于 $\bigotimes T$；并且，如果 c 的输出也独立于 $\{A_i\}$，那么它就独立于 $T \cup \{A_i\}$。因此，将 T 定义为所有满足输出独立于 $\{A_i\}$ 的输入 A_i 的并集，即得证。 □

E 定理90的证明

证明。令 $\big(G,\ W^k,\ P_*(O)\big)$ 为 id-cf 的输入，并令 D 为步骤 (2.) 之后的图，即表示所关注反事实的简化（未归一化）图。

假设步骤 (3.) 中继续执行的条件得到满足，即任意 U_X 的状态 λ_X 都已被吸收到相应的 c_X中。（注意，对于任意 $X \in R$，通过构造 D总能做到这一点，因为在 $\tilde{\rho}(G)$ 中，R 节点是根节点。）在这种情况下，对于任意 $X \in O \cup R$，D 中至多出现一次 c_X。不难看出，此时 D 必定是由 R-片段、复制映射，以及关于变量 $X \in O$的尖锐态和效应复合而成的。（参见图18a和公式57中的示例。）令 $\{F_1, ..., F_m\}$ 为 D的 R-片段集合，并回顾一下，R-片段对应于满足定义87中条件的极大片段。注意，D 的任意 R-片段的输入和输出导线类型均属于 O，根据定义，它们不可能属于 R 变量的类型。

假设算法没有输出失败，也就是说，对于作为片段 F_l 的导线类型出现的任何对象，步骤（4.1.1.a）或（4.1.1.b）以及（4.1.2.a）或（4.1.2.b）中的条件都满足。在这种情况下，式（58）和（59）中的相应重写显然会得到等价的图，因为它们只是 cd-范畴中的基本等价变换。

回想一下，定义87中 R-片段的定义并未规定关于机制 c_X for $X \in O$ 的连线条件，除了以下条件：在一个 R-片段中，一个机制的输出与另一个机制的输入之间的连通性，以及变量 $\tilde{}$in R 的状态与其的连线，都必须符合 $D_{\tilde{\rho}(G)}$。因此，重写后的片段 F_l（现在可能包含额外的复制映射）确实仍是一个 R-片段，并且对于任意变量 $X \in O$，至多只有一条导线从其 $\tilde{}$引出或接入其 F_l，绝不会两者兼有或多于一条。因此，步骤（4.2.）中指定的对象集满足：忽略必要的交换后，片段类型为 $\widetilde{F_l}$：$D_l \otimes F_l^{\text{in}} \to C_l \otimes F_{\text{输出}l}$；这里我们让这些符号同时表示对象集，以及按某种因子顺序排列的相应对象。

显然，这个片段现在由出现在带有网络图 $D_{\tilde{\rho}(G)}$ 的因果模型中的机制组成，而且其连接方式也

~ ~恰好如同 $D_{\tilde{\rho}(G)}$ 中那样——每个 F_l 也是 $D_{\tilde{\rho}(G)}$ 的一个片段。因此不难看出， F_l 与公式（60）中所示的信道 $P\big(C_l,\ F_{\text{输出}l};\text{do}(D_l),\text{do}(F_l^{\text{in}})\big)$ 完全相同。因此，步骤 (4.2.) 中的重写会产生一个等价的图。

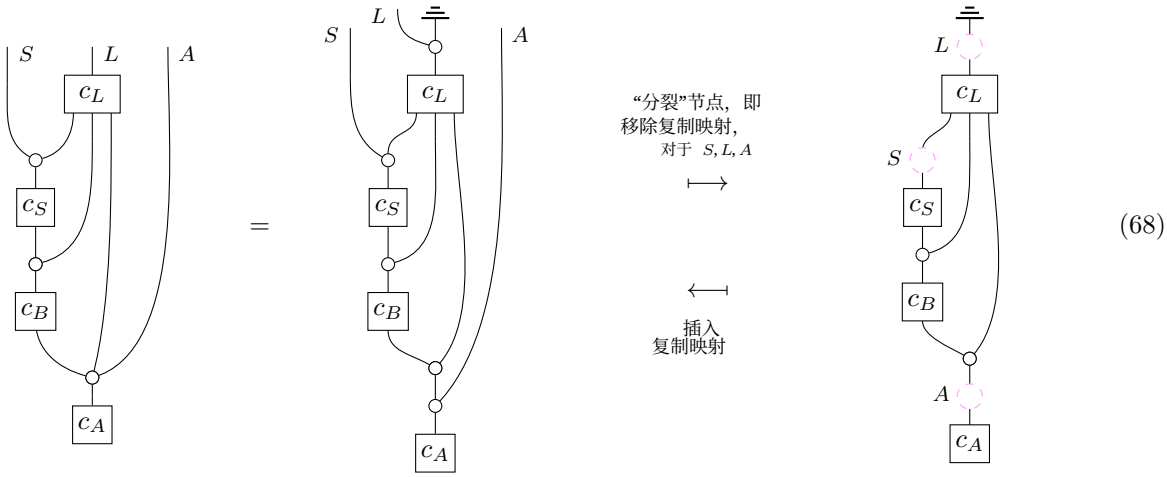
该图 D 现在是由尖锐态和效应、复制映射以及形如$P\big(C_l,\ F_l^{\text{out}}$的信道构成的复合；$\text{do}(D_l)$、$\text{do}(F_{\text{中}l})\big)$，其中后者位于 P_* 内，并由假设给定。总之，只要 id-cf 的输出不是失败，反事实 $C = norm(D)$就是可识别的，其表达式即为算法输出的表达式。 □

F 因果模型的分裂节点表示

除了用网络图表示因果模型之外，还有另一种同样具有图示性质的因果模型表示。受文献 [74], 的启发，我们将其称为因果模型的分裂节点表示。

基本上，给定一个具有输出变量 $O \subseteq V$、属于 C 的因果模型 $\mathbb{M}$，我们考察其网络图，但不再将每个变量 $X \in O$“复制出来”，而是断开或分裂从其机制 c_X引出的导线，从而创建一个可插入干预的“槽位”。这里我们考虑一种不同于上述定义的“干预”概念；它与非宽局部干预相关，但现在允许增加一条导线，用于记录干预的结果。

作为说明，我们回到贯穿全文的因果贝叶斯网络示例，其网络图如下左侧所示。对于所有变量 $X \in O$ $(= \{S, L, A\})$，分离出一个与 X 的输出导线对应的复制映射；也就是说，如果 $\text{Ch}(X) = \emptyset$，则像下方的 L 一样引入一个复制映射，并丢弃其中一条线；如果 $|\text{Ch}(X)| = 1$，则像下方的 S 一样不做处理；否则，像下方的 A 一样改写为两个复制映射。然后移除这些复制映射，使每个位置都留下两条悬空导线，从而形成槽位，如下方右侧粉色虚线圆圈所示。

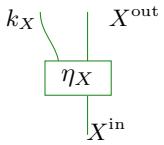


(68)

当然，也可以像上图右侧那样，以显而易见的方式，直接根据由机制（c_X）$_{X\in V}$和输出 $O \subseteq V$给出的因果模型，构造这种「带槽字符串图」。N

时, w hat do the sl槽位表示一个d what d整体图是t h总体 ll string di图 w ith slots 表示？ F或一个s l未标记为 X 时，进入该槽位并从中出来的两条导线是相应变量X的副本；为便于区分，可将它们标记为 X^{in} 和 X^{out}。槽位是可以插入干预的位置，而变量 X^{in} 和 X^{out} 分别表示潜在干预之前和之后的变量 X，因此将进入和离开槽位的变量这样标记。由此，槽位被理解为潜在干预位置。在槽位 X 处进行结果为 k_X 的干预，指的是一个

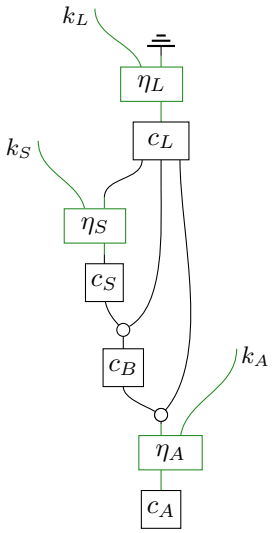
如下形式的信道 η_X：



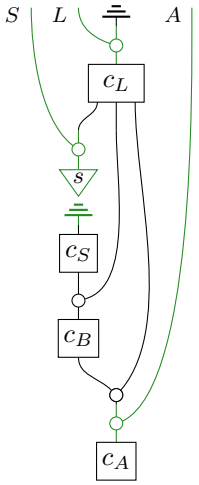
(69)

这种“干预”的概念（在形式上不属于定义 43）使我们能够对这样的情形建模：人们观测到X，但精度可能有限；或者某个主体抛掷一枚硬币，并根据结果将变量 X 制备为特定状态，同时 k_X 记录抛硬币的结果。

形式上，整体的带槽字符串图定义了一个高阶映射，也就是说，它是一个将某个范畴中的过程映射到同一范畴中过程的函数。因果模型的分裂节点表示将各个槽位上的干预选择映射为相应干预结果上的一个状态。图 19a 展示了这样得到的字符串图：将槽位 A、L 和S 上的干预选择代入式（68）右侧的图中。请注意，所得的图不是网络图，因为其中的方框 η_X 有多个输出，而这些方框不一定是复制映射。



(a)



(b)

图 19

尽管如此，信道 η_X 可以是复制映射，或者包含复制映射作为其组件之一，如图 19b 所示的特殊情形；该情形还原了例 44 中 do-干预 do($S = s$) 的示例。这里，η_A 和 η_L 各自都是复制映射，反映出 A 和 L 只是“观测到的”。在分裂节点模型的视角下，“观测” X 在形式上成为带有 $k_X = X$的干预的一个特例。

就如何表示因果模型的数据而言，看起来不过是对 $O \subseteq V$ 子集进行记账的另一种方式——复制映射与开放槽位——实际上定义了一个不同的整体对象，即高阶映射，同时也体现出一种倾向：人们心中已有特定的干预。关于高阶映射的形式化定义及更多技术细节——在无环因果模型这一语境下，它们总是构成所谓的“梳形映射”——请读者参阅相关文献（例如，参考文献 [113, 114, 115, 116,117, 118, 119]）。

注释 113.（定义高阶映射所需的范畴结构）：给定一个因果模型，将其表示为带槽字符串图总能得到定义良好的高阶映射，这一点仅由底层范畴的对称幺半结构保证——具体而言，无需假设紧闭性 *[119]*。

本研究聚焦于因果模型框架的某些方面，在这些方面采用高阶映射视角并无特别的启发性或益处。因此，我们将进一步探究其优势留待未来工作，并采用第3节所阐述的形式体系。