arXiv:2304.02023v1 [stat.ML] 4 Apr 2023

通过因果边缘问题界定因果概率

Numair Sani snumair1@jh.edu

Atalanti A. Mastakouri atalanti@amazon.de

Dominik Janzing janzind@amazon.de

摘要

因果概率在法律、医疗保健和公共政策的决策中发挥着基础性作用。然而，要对其进行点识别颇具挑战，需要单调性等强假设。若没有这类假设，现有研究需要观察多个包含相同治疗和结果变量的数据集，才能确定这些概率的界限。然而，在许多临床试验和公共政策评估案例中，存在相互独立的数据集，它们分别考察不同治疗对同一结果变量的影响。本文概述如何通过对基于此类独立数据集构建的结构因果模型施加反事实一致性约束（即“因果边缘问题”），显著收紧现有的因果概率界限。随后，我们介绍一种用于证伪反事实概率的新信息论方法，并使用条件互信息量化反事实影响。该方法可推广至任意离散变量和任意数量的治疗，并使因果边缘问题更易于解释。由于“界限是否足够紧”取决于用户判断，我们还提供了一种在界限不令人满意时进行推断的补充方法：一种基于最大熵的方法，为可能的结构因果模型空间定义度量，并提出在缺乏更多信息时用于推断反事实的熵最大化结构因果模型。

1 引言

文献 [1] 中定义的因果概率是一类因果量，在法律、公共政策和医疗保健领域 [2, 3] 均有应用。虽然在外生性等合理假设下，平均处理效应等因果参数可以根据观测数据分布点识别，但因果概率并非如此。具体而言，除了外生性之外，要实现点识别，还需要满足单调性 [4, 5]。然而，当这些假设缺乏依据，因而无法点识别 时，目标便改为部分识别，即利用观察性数据尽可能精确地界定这些概率。

关于部分识别及其在界定因果量中的应用，已有大量研究；例如包括[6, 7, 8, 9, 10]。因果概率的界定问题也已在 [7, 11, 12, 9, 10,13, 14, 15], 中得到探讨，不过，我们在此作出的贡献与现有研究有着根本区别。我们仅考虑图 1(a) 中所示的特定因果图，其中包含两个处理和一个按碰撞结构排列的单一结果。此外，我们假设处理和结果均为离散值；并且，我们无法获得两个处理 X 和 Y 与结果 Z的联合分布，只能获得该联合分布的边缘分布，即 $\mathbb{P}(X,\ Z)$和 $\mathbb{P}(Y,\ Z)$。虽然这是一个简单的例子，但它反映了现实世界中常见的情形：许多相互独立的研究关注同一结果，却研究不同的处理。随机对照试验（RCT）便是一个具体例子：相互竞争的制药公司针对同一种疾病测试不同的治疗方法。此时，联合分布指的是一种假设性的实验设计，其中两种处理分别独立随机分配。因此，尽管 [14, 11, 16, 4, 12, 10] 能够为因果概率等反事实量推导界限，但它们都假设能够获得所有相关变量的联合分布。相比之下，我们只假设能够获得相关变量联合分布的边缘分布。这种情形带来了额外挑战，因为通常无法根据边缘分布识别联合分布，因此难以应用现有界限。

为应对这些挑战，我们首先从信息论视角探讨如何在结合边缘分布时对反事实进行推理，并利用条件互信息给出关于反事实概率证伪的结果。接下来，我们利用 [17], 中展示的一致性约束，大幅收紧因果概率的现有界限。由于“足够紧”与否取决于方法使用者的判断，当界限不令人满意时，我们还提供另一种推断方法。这是一种基于最大熵的方法，用于选择“最佳”的结构因果模型（SCM）。尽管已有研究探讨了如何从数据中学习结构因果模型 [18, 19], ，但在我们的设定中，无法根据观测数据的边缘分布识别结构因果模型。在此设定下，我们应用最大熵原理 [20] ，学习一个与数据一致的结构因果模型，并以其熵作为证据的度量。最大熵的因果版本已在 [21] 中讨论，并在 [22]中用于结合数据集，但此前尚未用于从数据中推断结构因果模型。

我们的工作与 [23], 非常相似，但此处我们概述其主要区别：由于我们考虑的是特定的因果图结构，因此能够通过对相关线性规划进行符号求解，得出更多结果。此外，现有研究求得的是处理效应的界限，而我们能够给出符号界，从而收紧所有因果概率的现有界限。此外，我们还从信息论视角探讨了在存在额外变量时如何证伪结构因果模型，并对 [17] 的部分结果给出了简单解释。我们进一步提出一种基于最大熵的结构因果模型学习框架，用于选取符合观测边缘约束且熵最大的结构因果模型。

本文其余部分安排如下。第 2 节概述反事实与因果概率，第 3 节概述因果边缘问题 [17] ，并严格确立证伪结构因果模型与收紧因果概率界限之间的关系。随后，我们在第 4 节通过利用额外变量证伪结构因果模型，提出一种反事实推断的信息论方法。由于针对边缘问题的信息论界通常并不紧 [24], ，我们采用基于线性规划的方法，确立能够收紧因果概率界限的理论条件。我们将在第 5 节说明在何种假设下、如何以及能够将现有界限收紧多少。如果界限的紧致程度仍不理想，我们将在第 6 节提出一种基于最大熵的反事实推断方法。最后，我们在第 7 节给出总结。所有证明均见附录。

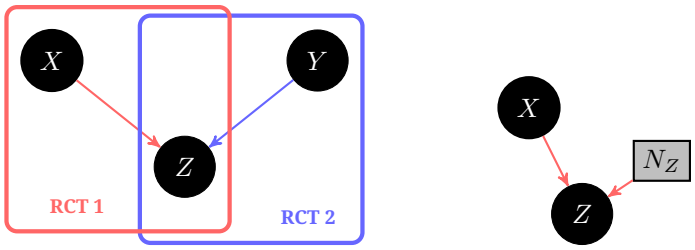
2 预备知识

本节简要介绍本文所用的相关概念，并为各主题的深入综述提供参考文献。我们首先定义因果模型，它构成了后续因果推理的基础。

文献中存在许多因果模型的表述形式，例如 [25] 提出的最精细的完全随机化因果可解释结构树图（FFRCISTG），以及 [26], 提出的不可知因果模型；本文采用 [1]中定义的结构因果模型（SCM）。形式上，SCM $\mathcal{M}$ 定义为一个元组 $\langle \mathbf{U}、\mathbf{V}\ \mathcal{F},\ \mathbb{P}\rangle$ ，其中 $\mathbf{U}$ 和$\mathbf{V}$ 分别表示外生随机变量和内生随机变量的集合。 $\mathcal{F}$ 表示一组函数，它们通过 $v \leftarrow f_V(pa_V,\ u_V)$ 确定 $V \in \mathbf{V}$ 的取值，其中 pa_V表示 V 的父节点， u_V表示与 V 相关的噪声变量的取值。 $\mathbb{P}$ 表示噪声变量 $\mathbf{U}$ 集合上的联合分布；由于假设噪声变量 $\mathbf{U}$ 相互独立，联合分布 $\mathbb{P}(\mathbf{U})$ 可分解为各噪声分布边缘分布的乘积。 $\mathcal{M}$ 在$\mathbf{V}$上诱导出一个观测数据分布，并与一个有向无环图（DAG） $\mathcal{G}$相关联。

定义结构因果模型（SCM）后，我们便可以定义子模型、潜在响应和反事实，具体定义见 [1]。给定因果模型 $\mathcal{M}$和随机变量$\mathbf{X} \subset \mathbf{V}$的一个实现值，子模型 $\mathcal{M}_x$ 对应于从$\mathcal{F}$ 中删除所有用于设定$\mathbf{X}$ 中元素取值的函数，并用常函数 $X = x$替换它们。该子模型刻画了干预 $do(X = x)$ 对 $\mathcal{M}$ 的影响。给定子集$\mathbf{Y} \subset \mathbf{V}$，潜在响应$\mathbf{Y}_x(u)$表示在外生变量$\mathbf{U}$取值为 u 时，满足 $\mathcal{M}_x$ 的 Y 取值。因此，反事实$\mathbf{Y}_x(u) = y$表示这样一种情形：若我们能够回溯事实，将 $X = x$设为 y，则潜在响应$\mathbf{Y}_x(u)$等于 y。当 u 由 $P(\mathbf{U})$ 生成时，我们便得到$\mathbf{Y}_x$ 对应的反事实随机变量，它们具有相应的概率分布。

一类重要的反事实概率在法律、医学和公共政策中有着广泛应用，称为因果概率 [1]。因果概率是一组与必要性与充分性概率（PNS）相关的五种反事实概率，定义如下。给定二元随机变量 Z 和 X，其中 Z 为结果，X 为暴露，令 z 表示随机变量 Z 取值为 1 的事件，令 z' 表示 Z



(a) (b)

图1：(a) **表示研究相同结果的独立临床试验的因果图**。 给定两个研究相同结果 Z、但采用两种不同治疗 X 和 Y的 RCT，结合这两项 RCT 比单独任一项 RCT 更有效地限制因果概率。(b) **SCM 的信息论视角** $X \to Z$。 在响应函数参数化下，噪声项 N_Z 可以视为世界其余部分对X 与 Z之间关系的影响

取值为 0，并对 x 和 x'采用类似的记法。于是，PNS 定义为

$$PNS \equiv \mathbb{P}(Z_x = z, Z_{x'} = z') \tag{1}$$

换言之，PNS 表示联合概率：反事实随机变量 Z_x 取值为 z，且反事实随机变量 $Z_{x'}$ 取值为 z'。在外生性条件下（定义为 $Z_x \perp\!\!\!\perp X$），其余因果概率，例如必要性概率（PN）和充分性概率（PS），均定义为 PNS 的函数（参见 [1]中的定理9.2.11）。因此，一旦 PNS 可识别，所有其他因果概率也都可以直接根据观测数据识别。

然而，要识别 PNS，我们必须作出诸如单调性 [7], 之类的假设，而在涉及实验药物、法律事务和职业健康的情境中，我们可能没有充分理由作出这些假设。尽管 PNS 不再是点识别的，但仍然可以利用部分识别文献中的工具对其进行有意义的界定。PNS 的一个重要界限定义于 [7], ，并在下文中给出，其中 p_{11}、 p_{10} 和 p_{00} 分别表示 $\mathbb{P}(Z = 1 \mid X = 1)$、$\mathbb{P}(Z = 1 \mid X = 0)$ 和 $\mathbb{P}(Z = 0 \mid X = 0)$。

$$\max[0, p_{11} - p_{10}] \leq PNS \leq \min[p_{11}, p_{00}] \tag{2}$$

在展示如何收紧因果概率的界限[1], 之前，我们需要介绍一些其他概念，包括 SCM 的响应函数参数化 [8]（称为 [27]中的 SCM 的规范参数化）以及因果边缘问题 [17], ，我们将在下一节中介绍这些内容。

3 反事实与因果边缘分布

本节首先简要概述因果边缘分布问题，该问题已在 [17]中得到研究。随后，我们将严格描述证伪结构因果模型与界定因果概率之间的关系。以下各节均假设因果图具有图1(b)所示的V结构。

3.1 因果边缘分布问题

定义结构因果模型时会引入一些自由度，即使对于反事实陈述而言，这些自由度也无关紧要（例如对噪声进行重新缩放）。真正重要的只是由 N_Z诱导出的函数 $f_Z(., N_Z)$。为消除这种无关的歧义，采用以下“响应函数”或“规范”表示法 [8, 27] 会很方便：

定义1（响应函数表示）若结构因果模型 $Z = f_Z(W, N_Z)$ 满足某些条件 $N_Z \perp\!\!\!\perp W$，其中 W、Z 是取值范围分别为有限集合 $\mathcal{W}$、$\mathcal{Z}$ 的变量，且 N_Z 的取值属于函数集合$\mathcal{W} \to \mathcal{Z}$，则称该结构因果模型具有“响应函数表示” $f_Z(W, n_Z) = n_Z(W)$。

这样， $P(N_Z)$ 是定义在函数上的分布。现在考察二元变量设定，我们有 $W = X$ 和 $W = (X, Y)$ 两种情况，其中 N_Z 分别取 4 个或 16 个值。在第一种情况下，这 4 个函数

[1]在我们的设定中，我们只考虑因果概率，因为文献中已对其进行了广泛研究。尽管如此，我们的理论也可直接应用于其他反事实概率，例如 $\mathbb{P}(Z_{x'} = z, Z_x = z')$，以及它的所有函数；我们将在附录第 18 节中讨论这一点。

读作 $\{\mathbf{0},\mathbf{1},$**恒等函数**、**逻辑取反**$\}$，其中前两个是常函数，而 **恒等函数** 和 **逻辑取反** 分别表示恒等函数和逻辑取反。令 $P(N_Z)$ 参数化为 $[a_0,a_1,a_2,a_3]$，其中每个 a_i 表示分配给 $\{\mathbf{0},\mathbf{1},$**恒等函数**、**逻辑取反**$\}$中相应函数的概率质量。

当 $W=X$ 时，观测分布 $\mathbb{P}(X,Z)$ 可用于对 $P(N_Z)$施加线性约束：通过强制结构因果模型与观测分布 $\mathbb{P}(X,Z)$ 满足一致性，可以得到两个约束；此外，利用概率分布之和为一这一性质，还可以得到一个约束。这样便留下一个自由参数，记为 λ_X，它能唯一地参数化下文所述的、由 $X\to Z$给出的结构因果模型。详细推导见附录第8节。

$$\begin{pmatrix}a_0\\a_1\\a_2\\a_3\end{pmatrix}=\begin{pmatrix}0\\1-p_{00}-p_{01}\\p_{00}\\p_{01}\end{pmatrix}+\lambda_X\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix}\tag{3}$$

其中 $\lambda_X\in[\lambda_X^{\min},\lambda_X^{\max}]$ 和 $\lambda_X^{\min}=$分别是 $\max[0,p_{00}+p_{01}-1]$ 和 $\min[p_{00},p_{01}]$。这里 $p_{00}\equiv\mathbb{P}(Z=0\mid X=$**0**$)$ 且 $p_{01}\equiv\mathbb{P}(Z=0\mid X=1)$。

[17] 指出，给定三个二元随机变量 X、Y 和 Z，且只能获得边缘分布$\mathbb{P}(X,Z)$ 和 $\mathbb{P}(Y,Z)$，当 $0<\mathbb{P}(X)<1$ 且 $0<\mathbb{P}(Y)<1$ 时，若因果图遵循图 1(a) 中的结构，则可对 λ_X 施加进一步的范围限制。具体做法是根据观测到的边缘分布，构造关于 $X\to Z$ 和 $Y\to Z$的结构因果模型，分别记为 $\mathcal{M}_X$ 和 $\mathcal{M}_Y$，然后强制 $\mathcal{M}_X$与 $\mathcal{M}_Y$满足反事实一致性。通过采用响应函数参数化，将关于 $\{X,Y\}\to Z$（记为 $\mathcal{M}$）的结构因果模型参数化为

$$X:=N_X\quad Y:=N_Y\quad Z:=h_{N_Z}(X,Y)\tag{4}$$

其中 N_Z 为 16 个可能函数 $h\colon\{X\times Y\}\to Z$ 编号。投影算子定义为

$$\mathcal{P}_x^X:h\to h(x,Y)=f(Y)\qquad\mathcal{P}_y^Y:h\to h(X,y)=f(X)\tag{5}$$

该投影算子使我们能够定义 $\mathcal{M}_X$ 与 $\mathcal{M}_Y$之间的反事实一致性约束，形式如下：

$$a_j(\lambda_X)=\sum_{y=0}^{1}\mathbb{P}(Y=y)\sum_{k=0}^{15}\mathbb{I}\left\{\mathcal{P}_y^Y(h_k)=f_j(X)\right\}c_k\tag{6}$$

$$b_j(\lambda_Y)=\sum_{x=0}^{1}\mathbb{P}(X=x)\sum_{k=0}^{15}\mathbb{I}\left\{\mathcal{P}_x^X(h_k)=f_j(Y)\right\}c_k\tag{7}$$

其中，c_k表示 $P(N_Z=k)$，即 $\mathcal{M}$和 $k\in0$、.....、15 中响应函数的权重；a_j和 b_j分别表示结构因果模型 $\mathcal{M}_X$ 和 $\mathcal{M}_Y$ 中响应函数的权重，且 $j\in\{0,1,2,3\}$。这些关系可简洁地表示为矩阵形式：

$$\mathbf{a}(\lambda_X)=\mathbf{Ac}\quad\mathbf{b}(\lambda_Y)=\mathbf{Bc}\tag{8}$$

A 和 **B** 的定义见 [17] 方程15。满足上述约束的 **c** 的取值集合构成一个凸多面体，我们将其记为 $\mathcal{C}$。该结论的证明见 [17],，我们将在附录第 15 节中介绍 $\mathcal{C}$。通过比较由 $\mathbb{P}(Z,\ X)$ 得到的 λ_X 范围与满足这些方程的 λ_X 范围，我们或许还能进一步限制 λ_X的范围。然而，尽管 [17] 展示了一个可以限制 λ_X 范围的例子，但要实现这种限制，$\mathbb{P}(X,\ Z)$ 和 $\mathbb{P}(Y,\ Z)$ 必须满足哪些条件，以及限制幅度如何，均未作讨论。本文将确立这些条件，并据此收紧因果概率的界限。为此，我们接下来将阐明 λ_X 与因果概率之间的联系。

3.2 将 λ_X 与反事实联系起来

在第 3.1 节中，我们看到了结构因果模型与 λ_X 之间的联系，以及限制 λ_X 的取值范围如何对应于对 $X\to Z$ 上结构因果模型的证伪。正如我们在附录第 9 节中所证明的，λ_X 也刻画了模型的反事实影响，即“在观察到 Z 和 X 的取值时，若 X 不同，Z 会不同的未归一化概率”，表示为 $p_{00}+p_{01}-2\lambda_X$。λ_X 的值越低，对应模型的反事实影响越强。因此，遵循 Pearl 的溯因、干预和预测三步过程 [1]，也可以用 λ_X 来表示因果概率。引理 1 确立了因果概率与 λ_X 之间的联系，并重新推导出 [7] 中关于 PNS 的界限。

引理 1（关联 λ_X 与 PNS）当 X 和 Z 是二元随机变量且 $X \to Z$ 时，$PNS = p_{00} - \lambda_X$。通过观察 $\mathbb{P}(X, Z)$ 推导出的 λ_X 的界限，重新得到 *[7]* 中关于 *PNS* 的界限。

$$\max[0, p_{11} - p_{10}] \leq p_{00} - \lambda_X \leq \min[p_{11}, p_{00}] \tag{9}$$

因此，收紧因果概率的界限，直接对应于证伪结构因果模型。对反事实施加一致性约束，是收紧因果概率界限的一种方法（我们将在第 5 节中对此进行严格探讨）；然而，当处理数量超过两个时，逐一列举相关约束会十分繁琐。正如我们将在下一节中展示的那样，信息论方法为在高维情形下（即提供大量处理变量，或变量可取许多不同值时）推理结构因果模型的证伪，提供了一个简单而优雅的框架，并带来额外的可解释性见解。

4 信息论反事实推断

本节将从信息论的角度探讨结构因果模型的证伪。为避免确定性关系中的互信息可能为无穷大等技术细节，我们将讨论范围限定在取值范围有限的变量上。令 X、Y、Z 分别取值于有限集合 $\mathcal{X}$、$\mathcal{Y}$、$\mathcal{Z}$。对于一族定义在 $X \to Z$ 上且与 $\mathbb{P}(X, Z)$ 一致的结构因果模型，互信息和熵等信息论度量为证伪结构因果模型提供了易于理解的视角。为说明这一点，我们将噪声 N_Z 视为一个随机变量，用以捕捉（并随后粗化）其余因果祖先对 Z 的 Z 的影响。我们简要回顾一下，对于两个随机变量 U、V，条件香农熵写作 $H(U|V) := -\sum_{u,v} \mathbb{P}(U = u, V = v) \log \mathbb{P}(U = u|V = v)$，条件互信息写作 $I(U\colon V \mid W) = H(U|W) + H(V|W) - H(U, V|W)$，见 [28]。

接下来，对于通过 $Z = f_Z(X, N_Z)$ 生成的观测边缘分布 $\mathbb{P}(X, Z)$，注意到 X 和 N_Z 完全决定了 $Z(H(Z \mid X, N_Z) = 0)$，因此 Z 的熵可以用互信息表示为

$$H(Z) = H(Z) - H(Z \mid X, N_Z) = I(X, N_Z : Z) = I(X : Z) + I(N_Z : Z \mid X) \tag{10}$$

对于固定的 $\mathbb{P}(Z, X)$，所有结构因果模型都具有相同的 $H(Z)$ 和相同的 $I(X\colon Z)$，因此它们也都具有相同的 $I(N_Z : Z \mid X)$，即在已知 X 的情况下噪声所包含的关于 Z 的信息。然而，我们还可以将 $I(X, N_Z\colon Z)$ 分解为

$$I(X, N_Z : Z) = I(N_Z : Z) + I(X : Z \mid N_Z). \tag{11}$$

在（11）中，不同结构因果模型之间相同的只有总和，而这两个项各不相同。这种分解揭示了结构因果模型之间一个可解释的特征差异：N_Z 是否已经提供了关于 Z 的一部分重要信息（即使不知道 X）。这里需要注意，（11）不一定指响应函数表述。相反，N_Z 可以是包含世界中所有 其他因素的任意噪声变量，这些因素与 X 一起决定 Z；此外，它还可能包含世界中与 Z 无关的部分（因为这些部分不会影响（11）中的各项）。以下结果给出了分解之间可能差异的范围：

引理 2（两个信息分量的取值范围）对于任意从 X 到 Z 的结构因果模型，我们有 e

$$I(N_Z : Z) \in [0, H(Z|X)]$$
$$I(X : Z \mid N_Z) \in [I(X : Z), H(X)].$$

引理 2 的证明见附录第 11 节。

例 1（独立二元变量）对于二元变量 X、Z，分解 (11) 在以下情形中最为明显：$P(Z = 1|X = 0) = P(Z = 1|X = 1) = 1/2$，此时两个极端结构因果模型分别对应于函数 **0**、**1** 的等权混合，或 **ID**、**NOT** 的等权混合。在第一种情况下，N_Z 完全决定了 Z，也就是说，我们有

$$I(N_Z : Z) = 1 \quad \textit{and} \quad I(X : Z \mid N_Z) = 0. \tag{12}$$

在第二种情况下，我们得到

$$I(N_Z : Z) = 0 \quad \textit{and} \quad I(X : Z \mid N_Z) = 1. \tag{13}$$

换言之，λ_X 的两个极端情形也对应于将信息$H(Z) = 1$ 分配到式(11)中两个项的两种极端情形。

要解释 $I(X:\ Z \mid N_Z)$，请注意，无条件互信息 $I(X:\ Z)$ 描述了 X在多大程度上控制总体均值中 Z 的统计特性。条件互信息 $I(X:\ Z \mid N_Z)$ 衡量反事实影响的强度，也就是说，它量化了在每个单独实例中（其中 N_Z 固定）， X 的变化在多大程度上会影响 Z 。[2] 对于 N_Z的固定值 n_Z ，条件互信息 $I(X:\ Z \mid N_Z = n_Z)$ 表示在固定响应函数下 X 与 Z 之间的互信息。因此，它也衡量对于每个满足 $N_Z = n_Z$的个体单位，当假设输入 X 不同时， Z 会如何变化。由此我们指出：**信息论度量能够捕捉 X 在调整所有其他影响 Z 的因素时，对 Z所产生的影响强度**。

值得注意的是，这种信息分解上的差异也是使用 $\mathbb{P}$（Y， Z）时结构因果模型$X \to Z$ 可证伪性的特征。请注意，不失一般性，我们可以假设 Y 是 N_Z 的函数，即$Y = g(N_Z)$ ，因为否则我们可以定义一个新的噪声变量 $N_Z := (N_Z,\ Y)$ ，以纳入此前未被 N_Z捕捉到的那部分世界。由于互信息对函数关系具有单调性，因此我们有

$$I(Y : Z) \le I(N_Z : Z). \tag{14}$$

因此，当 Y 包含的关于 Z 的信息多于假设的 N_Z 所预期时，任何假设的结构因果模型 $X \to Z$（其中 N_Z 与 Z 之间存在相应的信息关系）都会被 $\mathbb{P}$（Y， Z）证伪。因此，假设 $I(N_Z :\ Z)$取值较低的结构因果模型很容易被证伪；然而，如果 Y 包含的关于 Z 的信息远少于假设的 N_Z 所预期，这可能是因为 Y 未能向我们提供关于 N_Z相关部分的足够信息。

因此，按照 $I(X:\ Z \mid N_Z)$取值较大且 $I(N_Z:\ Z)$取值较低的定义，具有较大反事实影响的模型很容易被证伪。然而，当我们能够获得三变量观测并使用$P(X,\ Y,\ Z)$时，情况就有所不同。我们首先观察到：

引理3（条件影响的单调性）对于任意 Y 且其为 N_Z 的函数，我们有 *ve*

$$I(X : Z \mid Y) \le I(X : Z \mid N_Z).$$

[3]。与前面类似，只要我们观测到变量 Y，且对于假设的结构因果模型， $I(X:\ Z \mid Y)$的值高于 $I(X:\ Z \mid N_Z)$，就可以根据引理3证伪该结构因果模型。这也说得通，因为如上所述， $I(X:\ Z \mid N_Z)$的值较低，意味着在个体单位层面， X 对 Z 的影响不显著。如果我们发现由 Y的不同取值所界定的总体子群中， X 对 Z存在显著影响，这一点显然就会被证伪。

尽管上述信息论界很方便，但一般而言，它们只能提供关于 $\mathbb{P}(Y,\ Z)$ 与假设的结构因果模型一致性的必要 条件。[4] 然而，在二元 X， Z的情况下，项 $I(N_Z:\ Z)$ 和 $I(X:\ Z \mid N_Z)$ 已足以确定结构因果模型（就其响应函数表述而言）。

引理 4（将 λ_X 与信息联系起来）对于二元 X, Z，参数 λ_X 与 $I(N_Z :\ Z)$ 呈线性关系：

$$I(N_Z : Z) = H(Z) - H(X)\left[p_{00} + p_{01} - 2\lambda_X\right]. \tag{15}$$

[5]

因此， λ_X 的较大值对应于较低的反事实影响 $I(X:\ Z \mid N_Z)$。虽然信息论为利用高维数据证伪结构因果模型提供了优雅的框架，但它无法给出紧界。为获得紧界，在下一节中我们将展示如何施加反事实一致性约束。

5 因果概率的界

收紧因果概率的界可直接转化为一个由第3节的一致性约束导出的线性规划问题。为此，请注意， λ_X 的值对应于

[2]请注意，我们在这里并未混淆相关性与因果关系：先验而言，依赖强度并不衡量影响强度，但在此处这样做是合理的，因为 X 和 N_Z 是 Z仅有的原因。更明确地说，根据 [29] （参见其中的定理 4）， $I(X:\ Z \mid N_Z)$衡量图 1(b) 中 DAG 里的边 $X \to Z$ 的强度；而在仅由$X \to Z$ 构成的 DAG 中， $X \to Z$ 的强度由 $I(X:\ Z)$给出。

[3]证明见附录第 13 节

[4]基于香农熵次模性的信息论界（这些界限也并非紧确）已被推导出来，用于判定观测到的联合分布是否与假设的潜在因果结构 [24]一致。

[5]证明请参见附录第14节

列向量 $\mathbf{Ac}$。由此得到一个关于 $\mathbf{c}$ 的线性目标函数，而且如第3节所示，所有约束也都是关于 $\mathbf{c}$ 的线性函数。

如第 4 节所示，λ_X 的上界无法被证伪，因此我们只需求解最小化问题来界定 PNS。一般而言，即使只有两个边缘分布，线性规划也很难用符号形式求解。然而，当$\mathbb{P}(Y,\ Z)$具有退化条件分布时，例如 $\mathbb{P}(Z=0 \mid Y=0)=0$ 或 $\mathbb{P}(Z=0 \mid Y=0)=1$，我们可以给出 PNS 的符号界。

在随机化研究中可能会出现这类概率分布，例如疾病在不接受治疗时死亡率很高，或者治疗的效果极高。下面我们正式定义 $\mathbb{P}(Y,\ Z)$具有退化条件分布的含义。

定义2 如果以下任一条件成立，则称分布 $\mathbb{P}(Y,Z)$ 具有退化条件分布：$p'_{00}=0, p'_{01}=0, p'_{00}=1$或 $p'_{01}=1$。这里 $p'_{00}=\mathbb{P}(Z=0 \mid Y=0)$ 和 $p'_{01}=\mathbb{P}(Z=0 \mid Y=1)$。

对于 $\mathbb{P}(Y,\ Z)$ 具有退化条件分布的情形，我们给出相关结果，说明将此类 $\mathbb{P}(Y,\ Z)$ 与$\mathbb{P}(X,\ Z)$结合，如何收紧因果概率的界限。但首先，我们必须检查 $\mathbb{P}(X,\ Z)$ 和 $\mathbb{P}(Y,\ Z)$ 是否兼容——第3节中的一致性约束并未提供进行此项检查的条件。在下面的定理中，我们将证明，当 $\mathbb{P}(Y,\ Z)$ 具有退化条件分布时，可以通过检查一个简单条件来确保兼容性。

定理1 当 $\mathbb{P}(Y,Z)$ 具有退化条件分布时，除第*3*节中的一致性约束外，通过施加 $\mathbf{c}$ 以满足 $[\mathbf{Ac}]_0=\lambda_X^{\max}$ 所导出的条件，对于确保 $\mathbb{P}(X,Z)$ 与 $\mathbb{P}(Y,Z)$兼容而言，既是必要条件，也是充分条件。

下面我们给出 PNS 的收紧界限，并由此给出定理2中的因果概率界限。

定理2 当 $\mathbb{P}(Y,Z)$ 具有退化条件分布，且满足第*1*节所述条件时，施加反事实一致性会收紧 *PNS* 的上界，如下所示

$$\max[0, p_{11}-p_{10}] \leq PNS \leq \min\{p_{00}-\mathbb{D}_1, p_{11}-\mathbb{D}_0\} \tag{16}$$

其中 $\mathbb{D}_0 \equiv \mathbb{I}(p'_{\ 00}=0)\mathbb{P}(Y=0)+\mathbb{I}(p'_{01}=0)\mathbb{P}(Y=1)$ *and* $\mathbb{D}_1 \equiv \mathbb{I}(p'_{00}=1)\mathbb{P}(Y=0)+\mathbb{I}(p'_{01}=1)\mathbb{P})(Y=1)$

新的界限可能不够“紧”，有时我们必须依据不足的证据做出决策。下一节将介绍在这种情况下进行推理的方法。两个定理的证明分别见附录第16节和第17节。

6 使用最大熵对不精确概率进行推理

在许多情况下，因果概率的界限不具信息性，但仍必须做出决策。例如，负责开展临床试验的临床医生可能会考虑根据初步安全性数据提前终止试验；出于伦理原因，他们没有时间等待详尽结果再做决定。为减少患者伤害，临床医生必须依据证据不足的情况决定是否终止该试验。本节提出的理论旨在为这类证据不足的情形提供其他审慎决策方法。为在这种不确定性下进行推理，我们建议采用基于最大熵（MaxEnt）的方法。我们将介绍其细节、依据及不足，并由领域专家选择适当的框架。可以将这种方法视为对上述单纯形界限的补充，即当界限提供的信息不足，而我们仍不得不进行推理时使用。

与观测到的概率相容的、定义在 X、Y 和 Z 上的结构因果模型（对应于图 1(b) 中的因果图）构成一个凸集，其证明见 [17]。该凸集中的每个元素，即 $\mathbb{P}(N_Z)$，都是响应函数表述中 16 个可能函数上的一个分布。因此，该集合中存在唯一一个使 $H(N_Z)$ 最大化的 $\mathbb{P}(N_Z)$。我们以两种方式使用与熵最大化的 $\mathbb{P}(N_Z)$ 相对应的结构因果模型（称为最大熵结构因果模型）：首先，用它获得 λ_X 的点估计；其次，用它量化各种数据集 $\mathbb{P}(Z,\ Y)$ 的证据强度。

因果概率的点估计 λ_X： 使熵最大化的结构因果模型对应于 λ_X 的一个单一取值，可以将其视为一个“合理猜测”。这与最大熵的一般情形相同：在给定约束下使熵最大化的分布，可以被视为一个“合理猜测”，前提是必须假定分布唯一。关于

最大熵合理性的进一步讨论，请参阅 [30, 31, 32], ，以及 [33], ；在后者中，最大熵是通过博弈中最坏情形损失的最小化得出的。[6]

证据强度：当最大熵结构因果模型的熵较小时，可接受的结构因果模型空间也较小，因为大体积对应的最大熵分布较大（注意，反之则不成立[7]）。因此，与对P一无所知时相比，观察到 $P(Y,\ Z)$使最大熵结构因果模型的熵从 $X,\ Y$ 显著降低至 Z，这表明新增的观察显著缩小了可能的结构因果模型空间。从这个意义上说，熵的大幅降低也意味着，在观察到 $P(Y,\ Z)$之前可能存在的广大结构因果模型空间，如今将被证伪。

最大熵的这一性质促使我们将其用作度量，以量化不同数据集所提供的证据强度。我们的解决方案是根据熵降低的幅度，对各种可用数据集 $\mathbb{P}(Z,\ Y)$进行排序。这里，“强证据”意味着响应函数的不确定性（熵）显著降低，因此也意味着广大结构因果模型空间被证伪。当然，如果数据集 Y 几乎不包含关于 Z 的信息，那么它们也无法为我们提供多少关于 X 与 Z之间关系的信息。接下来，我们通过数值模拟展示最大熵的性质，并使用 Python 软件包 *scipy.optimize.minimize*[8] 求解该优化问题的对偶形式 [9]；其推导见 [34]。

以下所有图均针对固定的 $\mathbb{P}(X,\ Z)$ 生成，并改变 $\mathbb{P}(Y,\ Z)$。从图3(a)和图3(b)可以看出，最大熵结构因果模型的熵较低时，通常对应于 λ_X 取值范围较小的区域。更准确地说，从图3(c)可以看出，这种关系通常并非单调关系，因为正如预期，并非所有能够限制结构因果模型空间的 $\mathbb{P}(Z,\ Y)$ 都能限制 λ_X 的取值范围。请注意，图3(c)中的 $\lambda_X^{\max} - \lambda_X^{\min} = 0.2$ 是不受 $\mathbb{P}(Y,\ Z)$ 限制时区间 $[\lambda_X^{\max}, \lambda_X^{\min}]$ 的长度。对于 λ_X 的点估计，图2显示，最大熵所估计的 λ_X 位于 $[\lambda_X^{\max}, \lambda_X^{\min}]$的内部；它并非区间的中点，但在这里与中点相距不远。

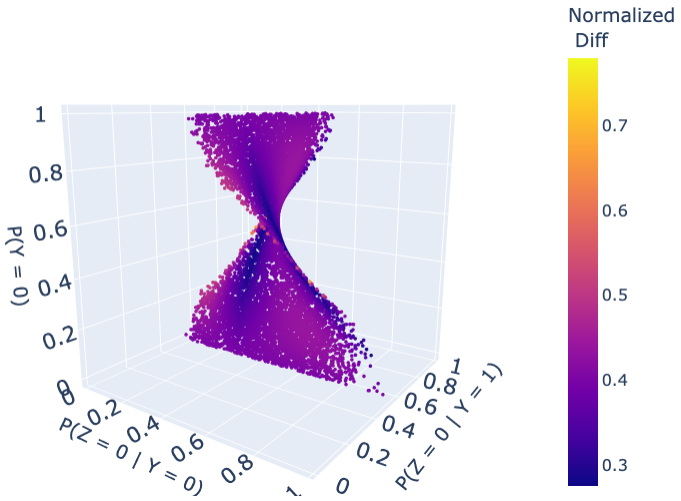


图2：**的最大熵估计**：最大熵估计的值 λ_X 与 $\lambda_X^{\max}$之差，再除以可能的 λ 范围。这里可以看到，最大熵估计的 λ_X 大多数时候都离区间 $[\lambda_X^{\min}, \lambda_X^{\max}]$的中点不远，对应的归一化差值为0.5。

[6]由于最大熵等价于最小化与均匀分布之间的KL距离，因此很容易将其推广到函数上的任意先验分布。对于具有许多输入的结构因果模型，我们可能希望偏好“简单”的函数，这与最大熵所依赖的、对所有函数隐含的均匀先验相冲突。[7]最大熵分布的较大取值并不意味着可行解空间很大。相反，其中可能只包含一个高熵分布[8]https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html[9]求解最大熵优化问题的对偶问题，可将待优化参数的数量从15个减少到4个（即约束的数量）

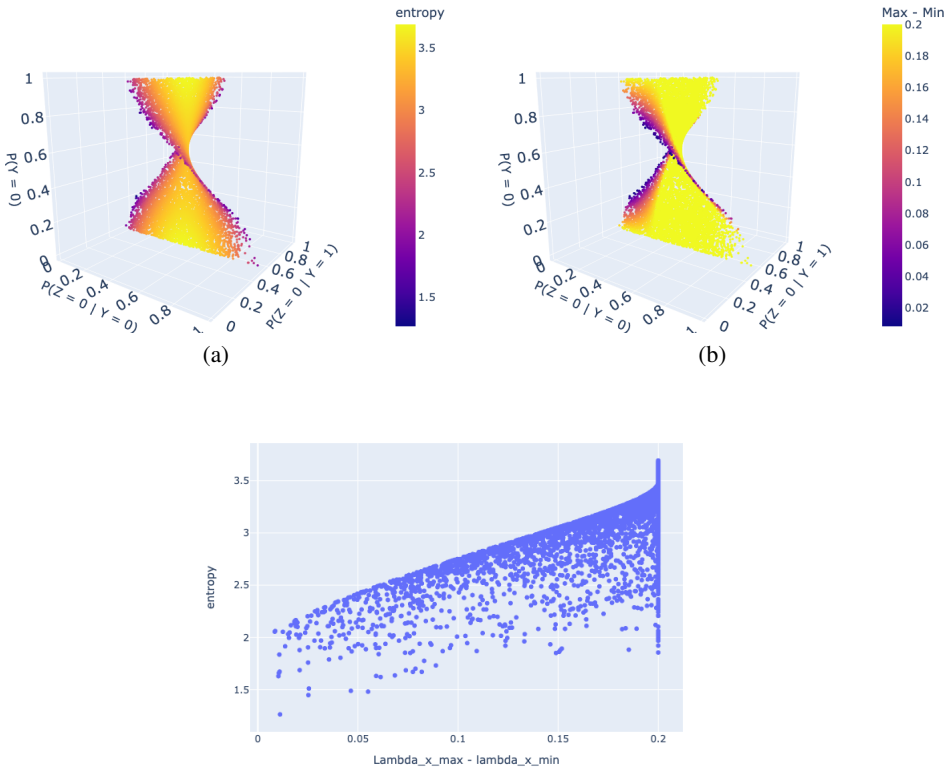


图 3：固定 p_{00}、 p_{01} 和 $\mathbb{P}(X=0)$ 的取值，改变 p'_{00}、 p'_{01} 和 $\mathbb{P}(Y=0)$，并绘制 (a) **作为证据强度的最大熵**：为所定义的结构因果模型族构建的最大熵结构因果模型的熵。使熵降低最多的 p'_{00}、 p'_{01} 和 $\mathbb{P}(Y=0)$ 取值被视为最强形式的证据。(b) $[\lambda_X^{\min}, \lambda_X^{\max}]$ 的**范围**，以及 (c) 联合结构因果模型的熵随区间 $[\lambda_X^{\min}, \lambda_X^{\max}]$ 的变化。

7 讨论

在介绍了 v 在反事实推断的各种方法中，我们简要强调几个关键讨论要点 s.

与数据融合的联系 一类类似的数据融合 研究 [35, 36] 旨在通过结合多个实验和观察数据集来识别干预效应。我们指出这类研究与本文工作的几个关键区别。首先，这种方法假设可以获取关于相同 一组随机变量的概率分布，只是对不同 变量进行干预。相比之下，我们假设只能获取关于不同 随机变量集合的概率分布，也就是说，结果 相同，但处理不同。其次，由于我们假设二元随机变量 X、 Y 和 Z 之间存在图 1(b) 所示的特定因果图，因此能够比上述研究给出更紧的因果概率界限。

二**元变量** 尽管我们将关注范围限定在二元变量上，但这种情况仍然涵盖了一个重要的现实场景。例如，类风湿关节炎这种疾病会采用多种药物进行治疗——巴瑞替尼 [37] 和非戈替尼 [38] 就是其中两种，且分别有独立的临床试验对其进行研究。这两项临床试验测量的是不同的治疗，但使用相同的结果指标（类风湿关节炎的ACR20应答标准[39]）。这是一个典型案例：我们的方法使临床医生能够利用所有这些独立的临床试验，强化反事实推断，并对药物疗效作出判断。

退化条件分布下的界限 要求边缘分布具有“退化条件分布”（即处于概率分布集合的边界，而我们的部分结果正是基于此条件）可能显得要求很高。然而，有许多临床试验的观测数据分布近似于具有退化条件分布的概率分布。例如，

BioNtech 针对 BNT162b2 疫苗开展的临床试验，其 [40]95% 的有效率可作为近似，用于利用我们收紧后的界限。

关于多个边缘数据集的未来工作如第 4 节前文所述，信息论视角也适用于包含高维随机变量不同子集的多个数据集的情形$\{X, Z, Y_1, ..., Y_d\}$。沿用 [24] 中基于香农熵次模性的思路，这些子集的熵和互信息构成了存在联合分布的必要条件；该联合分布应与假设的反事实影响 $I(X:\ Z \mid N_Z)$一致。这样，多个额外变量结合起来，可能会证伪从 X 到 Z 的结构因果模型（SCM）——这留待未来工作探讨。

结论 本文介绍了若干方法，通过合并处理变量不同而结果变量相同的数据集，获取有关因果概率的证据。为此，我们收紧了现有的因果概率界限，并提出了一种新的信息论视角，在证伪结构因果模型时兼具可解释性与可处理性优势。此外，在缺乏额外证据的情况下，我们提供了一个基于最大熵的反事实推断框架。我们的方法可推广到多个变量，也适用于因果概率之外的反事实陈述（Pearl 的因果阶梯第 3 阶）；相关说明见附录第 18 节。

附录

8 基于 $\mathbb{P}(Z,\ X)$的结构因果模型参数化

当 $X \to Z$、X 和 Z 均为二元随机变量时，其结构因果模型可用响应函数参数化表示为

$$X := N_X \tag{17}$$
$$Z := f_{N_Z}(X) \tag{18}$$

在响应函数参数化下，每个结构因果模型都可以由其噪声变量的分布唯一确定，$\mathbb{P}(N_Z,\ N_X)$。这里，N_Z是一个随机变量，其取值为函数集合 $\{\mathbf{0}, \mathbf{1}, \mathbf{ID},\ \mathbf{NOT}\}$中的某个函数值；而$N_X$是一个二元随机变量，直接生成 X的取值。虽然 $\mathbb{P}(N_X)$可以由 $\mathbb{P}(Z,\ X)$识别为$\mathbb{P}(X)$，但 N_Z 无法识别。不过，我们仍然可以利用 $\mathbb{P}(Z,\ X)$对 $\mathbb{P}(N_Z)$求界，如下所示。

设 $\mathbb{P}(N_Z)$参数化为 $[a_0, a_1, a_2, a_3]$，其中每个 a_i 表示分配给 $\{\mathbf{0}, \mathbf{1}, \mathbf{ID},\ \mathbf{NOT}\}$中相应函数的概率质量。由于当 $X = 0$ 时观测到 $Z = 0$ 的概率对应于分配给**0** 和**NOT** 函数的概率质量，因此可以用 $\mathbb{P}(Z,\ X)$约束 $\mathbb{P}(N_Z)$，如下所示

$$a_0 + a_2 = \mathbb{P}(Z = 0 \mid X = 0) \tag{19}$$
$$a_0 + a_3 = \mathbb{P}(Z = 0 \mid X = 1) \tag{20}$$
$$a_0 + a_1 + a_2 + a_3 = 1 \tag{21}$$

四个自由变量受到三个方程的约束，因此该方程组有一个自由参数。我们选择 a_0作为自由参数，并将其记为 λ_X。这样便可将上述方程改写为关于 λ_X 的表达式：

$$a_2 = \mathbb{P}(Z = 0 \mid X = 0) - \lambda_X \tag{22}$$
$$a_3 = \mathbb{P}(Z = 0 \mid X = 1) - \lambda_X \tag{23}$$
$$a_1 = 1 - \mathbb{P}(Z = 0 \mid X = 0) - \mathbb{P}(Z = 0 \mid X = 1) + \lambda_X \tag{24}$$

由于每个 a_i都必须是有效概率，即非负且小于或等于 **1**，因此我们可以得到 λ_X 的取值范围：

$$\max\{0, p_{00} + p_{01} - 1\} \le \lambda_X \le \min\{p_{00}, p_{01}\} \tag{25}$$

其中，$p_{00} = \mathbb{P}(Z = 0 \mid X = 0)$，且 $\mathbb{P}(Z = 0 \mid X = 1)$。与 $\mathbb{P}(Z,\ X)$ 相一致的任意结构因果模型的最终参数化为：

$$\mathbf{a}(\lambda_X) = \begin{pmatrix} 0 \\ 1 - p_{00} - p_{01} \\ p_{00} \\ p_{01} \end{pmatrix} + \lambda_X \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} \tag{26}$$

9 λX 捕捉反事实影响的合理性依据

反事实陈述“在观察到 Z 和 X 的特定取值时，如果 X 不同，Z 就会不同”可表示为概率 $\mathbb{P}(Z_{x'} = z' \mid Z = z,\ X = x)$。由于 **0** 和 **1** 函数无法根据不同输入产生不同输出，因此该反事实概率必定与赋予 **恒等** 和 **非** 函数的权重成正比；只有这两种函数能够在 X发生变化时改变输出值Z。因此，$\mathbb{P}(Z_{x'} = z' \mid Z = z,\ X = x)$ $\propto p_{00} + p_{01} - 2\lambda_X$。$p_{00}$和 p_{01}都是由 $\mathbb{P}(Z,\ X)$ 确定的函数，因此，与 $\mathbb{P}(Z,\ X)$ 相一致的结构因果模型的反事实影响仅随 λ_X变化。

10 引理 1 的证明

PNS 表示两个事件 $Z_x = z$ 和 $Z_{x'} = z'$的联合概率，即 $\mathbb{P}(Z_x = z,\ Z_{x'} = z')$，其中$z = 1$、$z' = 0$，对于 x 和 x'同样成立。因此，PNS 对应于赋予 **ID**函数的概率质量，因为只有该函数能够在 $X = 1$ 时输出 $Z = 1$，并在 $X = 0$时输出 $Z = 0$。根据第8节中介绍的 $\mathcal{M}_X$ 参数化，PNS 将等于 $p_{00} - \lambda_X$。

现在，通过代入第8节中介绍的 λ_X 界限，可以得到 PNS 的界限。 λ_X 的下界将转化为 PNS 的上界，反之亦然。由此得到界限

$$\max[0, p_{11} - p_{10}] \leq PNS \leq \min[p_{00}, p_{11}] \tag{27}$$

由此可恢复 [7]关于 PNS 的界限。

11 引理 2 的证明

为了证明这个引理，我们首先陈述并证明以下引理。

引理 5 对于任意三个随机变量 U、V、W，若满足 $U \perp\!\!\!\perp V$，则有

$$I(U:W) \leq I(U:W \mid V). \tag{28}$$

证明：利用条件互信息的性质，$I(U:W \mid V)$ 可以写成

$$I(U:W \mid V) = H(U \mid V) - H(U \mid V, W) \tag{29}$$

由于 $U \perp\!\!\!\perp V$，我们可以将上式改写为

$$I(U:W \mid V) = H(U) - H(U \mid V, W) \tag{30}$$

$$\tag{31}$$

类似地，可以将 $I(U:W)$ 改写为

$$I(U:W) = H(U) - H(U \mid W) \tag{32}$$

由于 $H(U \mid V, W) \leq H(U \mid W)$，引理成立。
以上述引理的表述为基础，我们现在可以开始证明。

引理 3 2的证明：我们首先给出 $I(X:Z \mid N_Z)$的界限。$I(X:Z \mid N_Z)$ 的下界由 X 与 N_Z 的独立性以及引理 5 得出。$I(X:Z \mid N_Z)$ 的上界如下所示。

$$I(X:Z \mid N_Z) = H(X|N_Z) - H(X|N_Z, Z) \leq H(X|N_Z) \leq H(X) \tag{33}$$

因此，$I(X:Z \mid N_Z)$ 的界限为

$$I(X:Z) \leq I(X:Z \mid N_Z) \leq H(X) \tag{34}$$

为说明 $I(N_Z:Z)$ 的界限，我们注意到，下界由互信息的非负性得到，并在 $N_Z \perp\!\!\!\perp Z$时取到。对于上界，注意

$$I(N_Z:Z) = H(Z) - I(X:Z \mid N_Z) \leq H(Z) - I(X:Z) = H(Z|X) \tag{35}$$

$$I(N_Z:Z) \leq H(Z \mid X) \tag{36}$$

$$0 \leq I(N_Z:Z) \leq H(Z \mid X) \tag{37}$$

12 例 1 的推导

为推导出所展示的结果，我们首先注意到 $\mathbb{P}(Z=0 \mid X=0) = \mathbb{P}(Z=0 \mid X=1) = \frac{1}{2}$ 蕴含$\mathbb{P}(Z=0) = \frac{1}{2}$，因此 $H(Z) = 1$。接下来，$\mathbb{P}(Z=0 \mid X=0) = \mathbb{P}(Z=0 \mid X=1) = \frac{1}{2}$，也将 λ_X 的取值范围设定为 $\lambda_X \in [0, \frac{1}{2}]$。上界 $\lambda_X = \frac{1}{2}$ 对应于反事实影响为 0 的模型，而该模型对0 函数和1 函数赋予相同的权重。因此，$\mathbb{P}(N_Z)$ 可写为

$$\mathbb{P}(N_Z = 0) = \frac{1}{2} \tag{38}$$

$$\mathbb{P}(N_Z = 1) = \frac{1}{2} \tag{39}$$

$$\mathbb{P}(N_Z = 2) = 0 \tag{40}$$

$$\mathbb{P}(N_Z = 3) = 0 \tag{41}$$

给定上述 $\mathbb{P}(N_Z)$ 的分布，我们可以计算 $\mathbb{P}(Z,\ N_Z)$ 的分布，如下所示

$$\mathbb{P}(Z=0,N_Z=0)=\frac{1}{2} \tag{42}$$
$$\mathbb{P}(Z=0,N_Z=1)=0 \tag{43}$$
$$\mathbb{P}(Z=0,N_Z=2)=0 \tag{44}$$
$$\mathbb{P}(Z=0,N_Z=3)=0 \tag{45}$$
$$\mathbb{P}(Z=1,N_Z=0)=0 \tag{46}$$
$$\mathbb{P}(Z=1,N_Z=1)=\frac{1}{2} \tag{47}$$
$$\mathbb{P}(Z=1,N_Z=2)=0 \tag{48}$$
$$\mathbb{P}(Z=1,N_Z=3)=0 \tag{49}$$

利用计算得到的分布，$I(N_Z:\ Z)$ 的公式给出 $I(N_Z:\ Z)=1$。再利用分解和 $H(Z)=1$，即可证明在满足 $\lambda_X=\frac{1}{2}$ 的结构因果模型中，$I(X:\ Z\mid N_Z)=0$。

对于满足 $\lambda_X=0$ 的结构因果模型，$\mathbb{P}(N_Z)$ 会对 **ID** 函数和 **NOT** 函数赋予相同的权重。因此，$\mathbb{P}(Z,N_Z)$ 将等于

$$\mathbb{P}(Z=0,N_Z=0)=0 \tag{50}$$
$$\mathbb{P}(Z=0,N_Z=1)=0 \tag{51}$$
$$\mathbb{P}(Z=0,N_Z=2)=\frac{1}{2}\mathbb{P}(X=0)=\frac{1}{4} \tag{52}$$
$$\mathbb{P}(Z=0,N_Z=3)=\frac{1}{2}\mathbb{P}(X=1)=\frac{1}{4} \tag{53}$$
$$\mathbb{P}(Z=1,N_Z=0)=0 \tag{54}$$
$$\mathbb{P}(Z=1,N_Z=1)=0 \tag{55}$$
$$\mathbb{P}(Z=1,N_Z=2)=\frac{1}{2}\mathbb{P}(X=1)=\frac{1}{4} \tag{56}$$
$$\mathbb{P}(Z=1,N_Z=3)=\frac{1}{2}\mathbb{P}(X=0)=\frac{1}{4} \tag{57}$$

因此，$I(N_Z:\ Z)=0$，并且根据分解，$I(X:\ Z\mid N_Z)=1$。

13 引理 3 的证明

由于 $X\perp\!\!\!\perp N_Z$ 和 $Y=g(N_Z)$，我们还有 $X\perp\!\!\!\perp(Y,N_Z)$。应用弱并图拟阵公理可得 $X\perp\!\!\!\perp N_Z\mid Y$。将引理 5 应用于变量 X,N_Z,Z，并给定条件分布 Y，可得 $I(X:\ Z\mid Y)\le I(X:\ Z\mid Y,N_Z)=I(X:\ Z\mid N_Z)$。

14 引理 4 的证明

根据互信息的定义，我们有 $I(N:\ Z)=-H(Z\mid N_Z)+H(Z)$。

给定 $\mathbb{P}(Z,\ X)$，通过边缘化可知 $\mathbb{P}(Z)$，因此它对每个结构因果模型都是已知且固定的。接下来我们需要计算 $H(Z\mid N_Z)$，为此需要计算 $\mathbb{P}(Z\mid N_Z)$ 和 $\mathbb{P}(Z,\ N_Z)$，具体列举如下

对于 $N_Z=0$

$$\mathbb{P}(Z=0\mid N_Z=0)=1 \tag{58}$$
$$\mathbb{P}(Z=1\mid N_Z=0)=0 \tag{59}$$
$$\mathbb{P}(Z=0,N_Z=0)=\lambda_X \tag{60}$$
$$\mathbb{P}(Z=1,N_Z=0)=0 \tag{61}$$

对于 $N_Z = 1$

$$\mathbb{P}(Z = 0 \mid N_Z = 1) = 0 \tag{62}$$
$$\mathbb{P}(Z = 1 \mid N_Z = 1) = 1 \tag{63}$$
$$\mathbb{P}(Z = 0, N_Z = 1) = 0 \tag{64}$$
$$\mathbb{P}(Z = 1, N_Z = 1) = 1 - p_{00} - p_{01} + \lambda_X \tag{65}$$

对于 $N_Z = 2$

$$\mathbb{P}(Z = 0 \mid N_Z = 2) = \mathbb{P}(X = 0) \tag{66}$$
$$\mathbb{P}(Z = 1 \mid N_Z = 2) = \mathbb{P}(X = 1) \tag{67}$$
$$\mathbb{P}(Z = 0, \ N_Z = 2) = \mathbb{P}(X = 0) \times (p_{00} - \lambda_X) \tag{68}$$
$$\mathbb{P}(Z = 1, N_Z = 2) = \mathbb{P}(X = 1) \times (p_{00} - \lambda_X) \tag{69}$$

对于 $N_Z = 3$

$$\mathbb{P}(Z = 0 \mid N_Z = 3) = \mathbb{P}(X = 1) \tag{70}$$
$$\mathbb{P}(Z = 1 \mid N_Z = 3) = \mathbb{P}(X = 0) \tag{71}$$
$$\mathbb{P}(Z = 0, N_Z = 3) = \mathbb{P}(X = 1) \times (p_{01} - \lambda_X) \tag{72}$$
$$\mathbb{P}(Z = 1, N_Z = 3) = \mathbb{P}(X = 0) \times (p_{01} - \lambda_X) \tag{73}$$
$$\tag{74}$$

利用这一点计算 $H(Z \mid N_Z)$，注意到含有 $N_Z = 0$ 和 $N_Z = 1$ 的项在$H(Z \mid N_Z)$的表达式中会相互抵消，剩下

$$H(Z \mid N_Z) = -\mathbb{P}(X = 0) \times (p_{00} - \lambda_X) \log \mathbb{P}(X = 0) - \mathbb{P}(X = 1) \times (p_{00} - \lambda_X) \log \mathbb{P}(X = 1) \tag{75}$$
$$- \mathbb{P}(X = 1) \times (p_{01} - \lambda_X) \log \mathbb{P}(X = 1) - \mathbb{P}(X = 0) \times (p_{01} - \lambda_X) \log \mathbb{P}(X = 0) \tag{76}$$
$$= H(X)[p_{00} - p_{01} - 2\lambda_X] \tag{77}$$

将其代入 $I(N_Z : Z)$的表达式

$$I(N_Z : Z) = -H(X) \left[p_{00} + p_{01} - 2\lambda_X\right] + H(Z) \tag{78}$$

关于凸多面体的 15 个细节 $\mathcal{C}$

此处我们给出描述 $\mathcal{C}$ 的等式和不等式约束。 $\mathcal{C}$ 是满足以下约束的 $\mathbf{c}$ 的取值集合，其推导过程见 [17]。

$$\tilde{\mathbf{A}}\mathbf{c} \preceq \tilde{\mathbf{b}} \tag{79}$$
$$\tilde{\mathbf{C}}\mathbf{c} = \tilde{\mathbf{d}} \tag{80}$$

~ ~ ~ ~我们在下文定义矩阵$\mathbf{A}$、$\mathbf{b}$、$\mathbf{C}$ 和$\mathbf{d}$。记号$\mathbf{a} \preceq \mathbf{b}$表示对所有元素均成立的 $a_i \leq b_i$。

不等式约束 c ~ ~**约束**。 此处我们使用 [17] 方程15. 中定义的$\mathbf{A}$ 和$\mathbf{B}$， 而$\mathbf{A}$ 和$\mathbf{b}$ 定义如下：

$$\tilde{\mathbf{A}} := \begin{pmatrix} \mathbf{A} \\ -\mathbf{A} \\ \mathbf{B} \\ -\mathbf{B} \\ -\mathbb{I} \end{pmatrix}, \qquad \tilde{\mathbf{b}} := \begin{pmatrix} \tilde{\mathbf{a}}_1 \\ \tilde{\mathbf{a}}_2 \\ \tilde{\mathbf{b}}_1 \\ \tilde{\mathbf{b}}_2 \\ \mathbf{0} \end{pmatrix} \tag{81}$$

~ ~为定义 ~$\mathbf{a}_1$, ~$\mathbf{a}_2$, $\mathbf{b}_1$, $\mathbf{b}_2$， 我们先在下文定义$\mathbf{a}_0$ 和 $\mathbf{b}_0$

$$\mathbf{a}_0 = \begin{pmatrix} 0 \\ 1 - P(Z = 0|X = 0) - P(Z = 0|X = 1) \\ P(Z = 0|X = 0) \\ P(Z = 0|X = 1) \end{pmatrix} \tag{82}$$

$$\mathbf{b}_0 = \begin{pmatrix} 0 \\ 1 - P(Z = 0|Y = 0) - P(Z = 0|Y = 1) \\ P(Z = 0|Y = 0) \\ P(Z = 0|Y = 1) \end{pmatrix} \tag{83}$$
$$\tag{84}$$

˜ ˜现在， ˜$\mathbf{a}_1$, ˜$\mathbf{a}_2$,$\mathbf{b}_1$, $\mathbf{b}_2$ 定义如下:

$$\tilde{\mathbf{a}}_1 = \mathbf{a}_0 + \begin{pmatrix} \lambda_X^{\max} \\ \lambda_X^{\max} \\ -\lambda_X^{\min} \\ -\lambda_X^{\min} \end{pmatrix} \quad \tilde{\mathbf{a}}_2 = -\mathbf{a}_0 + \begin{pmatrix} -\lambda_X^{\min} \\ -\lambda_X^{\min} \\ \lambda_X^{\max} \\ \lambda_X^{\max} \end{pmatrix} \tag{85}$$

$$\tilde{\mathbf{b}}_1 = \mathbf{b}_0 + \begin{pmatrix} \lambda_Y^{\max} \\ \lambda_Y^{\max} \\ -\lambda_Y^{\min} \\ -\lambda_Y^{\min} \end{pmatrix} \quad \tilde{\mathbf{b}}_2 = -\mathbf{b}_0 + \begin{pmatrix} -\lambda_Y^{\min} \\ -\lambda_Y^{\min} \\ \lambda_Y^{\max} \\ \lambda_Y^{\max} \end{pmatrix} \tag{86}$$

并且$\mathbf{I}$ 是一个包含 16 个 1 的向量，而$\mathbf{0}$ 是一个由 16 个 0 组成的向量。

˜ ˜**等式约束** 为定义等式约束，我们将$\mathbf{C}$ 和$\mathbf{d}$ 定义为

$$\tilde{\mathbf{C}} := \begin{pmatrix} \tilde{\mathbf{C}}_{\mathbf{A}} \\ \tilde{\mathbf{C}}_{\mathbf{B}} \\ \tilde{\mathbf{C}}_{\mathbf{1}} \end{pmatrix} \qquad \tilde{\mathbf{d}} := \begin{pmatrix} \tilde{\mathbf{d}}_{\mathbf{a}_0} \\ \tilde{\mathbf{d}}_{\mathbf{b}_0} \\ 1 \end{pmatrix} \tag{87}$$

其中

$$\tilde{\mathbf{C}}_{\mathbf{A}} := \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \mathbf{A}, \quad \tilde{\mathbf{d}}_{\mathbf{a}_0} := \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \mathbf{a}0 \tag{88}$$

$$\tilde{\mathbf{C}}_{\mathbf{B}} := \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \mathbf{B} \quad \tilde{\mathbf{d}}_{\mathbf{b}_0} := \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \mathbf{b}_0 \tag{89}$$

˜并且$\mathbf{C_1}$ 是一个由 16 个 1 组成的行向量。

16 定理1的证明

除重新参数化外，本证明还使用了一个引理和一个命题。我们先陈述并证明它们，再用它们证明定理1。首先，我们介绍一个引理，说明条件分布退化的概率分布与一类特殊结构因果模型之间的联系。

引理6 分布 族 $\mathbb{P}(Y, Z)$ 中，退化条件分布对应于满足 $\lambda_Y^{\min} = \lambda_Y^{\max}$ 的结构因果模型族。

证明 下面按情况逐一证明，其中 p'_{00} 表示 $\mathbb{P}(Z = 0 \mid Y = 0)$ 和 $p'_{01} = \mathbb{P}(Z = 0 \mid Y = 1)$:

• 当 $p'_{00} = 0$ 时， $\lambda_{\text{最大值}_Y} = 0$。 $p'_{00} + p'_{01} - 1$ 至多等于 0。因此 $\lambda_{\text{最小值}_Y} = 0$。所以 $\lambda_{\text{最小值}Y} = \lambda_{\text{最大值}_Y} = 0$。
• 同理，当 $p'_{01} = 0$ 时， $\lambda_{\text{最大值}_Y} = 0$。 $p'_{00} - 1$ 至多等于 0。因此 $\lambda_{\text{最小值}_Y} = 0$。所以 $\lambda_{\text{最小值}Y} = \lambda_{\text{最大值}_Y} = 0$。
• 当 $p'_{00} = 1$ 时， $\lambda_{\text{最小值}Y} =$最大值$\{p'_{01}, 0\}$，且 $\lambda_{\text{最大值}Y} =$最小值$\{p'_{01}, 1\}$。由于 p'_{01} 是概率且非负，$\lambda_{\text{最小值}Y} = \lambda_{\text{最大值}_Y} = p'_{01}$。
• 类似的论证表明，当 $p'_{01} = 1$时， $\lambda_{\text{最小值}Y} = \lambda_{\text{最大值}_Y} = p'_{00}$。
接下来，我们陈述命题 4.1 的一个改述版本，见 [17]。

命题 1 若 $\mathcal{C}$ 非空，则 $\exists$**存在** $\in \mathcal{C}$ 使得 $[A\mathbf{c}]_0 = \lambda_X^{\max}$且 $[B\mathbf{c}]_0 = \lambda_Y^{\max}$。

有了这些结果，我们就可以证明定理1了。

定理1的证明 要证明条件 A 对于 B是充要的，我们必须证明 $A \implies B$ 以及 $B \implies A$。在本例中，这意味着我们必须证明以下两个陈述:
• 当 $\mathcal{C}$ 非空时，存在 $\exists\mathbf{c}$，它是一个有效的概率分布，满足对 $\mathcal{C}$ 施加的所有约束，并且具有$[\mathbf{Ac}]_0 = \lambda_{\max X}$。

• 当结合多面体约束，由要求$\mathbf{c}$ 满足 $[\mathbf{Ac}]_0 = \lambda_{\max X}$ 所推导出的条件成立时， $\mathcal{C}$ 非空。

注意， [17] 中的命题 4.1 证明了第一个陈述；对于其余证明，我们将重点放在第二个陈述上。证明针对退化条件分布中出现的四种情况（见 6）逐一展开，但我们仅给出 $\mathbb{P}(Z=0 \mid Y=1) = 0$的证明，其余情况也可以采用类似的推导。

16.1 $\mathbb{P}(Z=0 \mid Y=1) = 0$

当 $\mathbb{P}(Z=0 \mid Y=1) = 0$、 $\lambda_Y^{\max} = \lambda_Y^{\min} = 0$，且结构因果模型 $Y \to Z$ 可参数化如下，其中 p'_{00} 表示 $\mathbb{P}(Z=0 \mid Y=0)$ 和 $p'_{01} = \mathbb{P}(Z=0 \mid Y=1)$。注意，由于 $\lambda_Y^{\max} = \lambda_Y^{\min} = 0$，$\mathbf{Bc}$ 上的所有不等式约束现在都是等式约束。

$$\mathbf{b}(\lambda_Y) = \begin{pmatrix} 0 \\ 1 - p'_{00} \\ p'_{00} \\ 0 \end{pmatrix} \tag{90}$$

因此，约束 $\mathbf{b}(\lambda_Y) = \mathbf{Bc}$ ，结合 $c_i \in \mathbf{c}$ 的非负性，以及 $\mathbb{P}(Y=0)$ 和 $\mathbb{P}(X=0)$ 均与 0 和 1 保持正距离，迫使$\mathbf{c}$ 中的以下元素等于 0：

$$c_0 = c_1 = c_2 = c_3 = c_4 = c_5 = c_6 = c_7 = c_8 = c_9 = c_{12} = c_{13} = 0 \tag{91}$$

将$\mathbf{c}$ 的相关值设为零，并将$\mathbf{c}$ 上的其余等式约束化为简化行阶梯形（RREF）后，我们得到下列方程组。

$$c_{10} - c_{15} = \frac{p_{00} + p_{01} - \mathbb{P}(Y=0)}{\mathbb{P}(Y=0)} \tag{92}$$

$$c_{11} + c_{15} = \frac{\mathbb{P}(Y=0) - p_{00}}{\mathbb{P}(Y=0)} \tag{93}$$

$$c_{14} + c_{15} = \frac{\mathbb{P}(Y=0) - p_{01}}{\mathbb{P}(Y=0)} \tag{94}$$

由于有三个方程和四个未知数，这个方程组有无穷多解。然而，以 $[A\mathbf{c}]_0 = \lambda_X^{\max}$ 的形式给出的附加约束提供了第四个方程，从而唯一确定该方程组；求解后得到以下 c_i 的取值。

$$[A\mathbf{c}]_0 = \lambda_X^{\max} \implies \mathbb{P}(Y=0)c_{10} = \lambda_X^{\max} \implies c_{10} = \frac{\lambda_X^{\max}}{\mathbb{P}(Y=0)} \tag{95}$$

$$c_{15} = \frac{\lambda_X^{\max} - p_{00} - p_{01} + \mathbb{P}(Y=0)}{\mathbb{P}(Y=0)} \tag{96}$$

$$c_{11} = \frac{p_{01} - \lambda_X^{\max}}{\mathbb{P}(Y=0)} \tag{97}$$

$$c_{14} = \frac{p_{00} - \lambda_X^{\max}}{\mathbb{P}(Y=0)} \tag{98}$$

在找到唯一 满足上述方程组的 $\mathbf{c}$ 值后，我们列出使其成为有效概率所需的条件的条件 ，从而得到以下约束。

$$0 \le \lambda_X^{\max} \le \mathbb{P}(Y=0) \tag{99}$$

$$\lambda_X^{\max} \le p_{01} \tag{100}$$

$$p_{01} - \lambda_X^{\max} \le \mathbb{P}(Y=0)\lambda_X^{\max} \le p_{00} \tag{101}$$

$$p_{00} - \lambda_X^{\max} \le \mathbb{P}(Y=0) \tag{102}$$

$$\lambda_X^{\max} \ge p_{00} + p_{01} - \mathbb{P}(Y=0) \tag{103}$$

$$\lambda_X^{\max} \le p_{00} + p_{01} \tag{104}$$

由这些结果　约束中，有些由定义隐含，因此我们将其剔除，留下以下　g

$$\lambda_X^{\max} \le \mathbb{P}(Y=0) \tag{105}$$
$$p_{01} - \lambda_X^{\max} \le \mathbb{P}(Y=0) \tag{106}$$
$$p_{00} - \lambda_X^{\max} \le \mathbb{P}(Y=0) \tag{107}$$
$$\lambda_X^{\max} \ge p_{00} + p_{01} - \mathbb{P}(Y=0) \tag{108}$$
$$\tag{109}$$

现在我们证明，当这些条件成立时，$\mathcal{C}$ 中的其余约束也成立，因此我们找到了 $\mathcal{C}$中的一个点，证明它非空。接下来，将 c_{10}、 c_{11}、 c_{14}和 c_{15} 的值以及上述推导出的条件代入，检查 $\mathcal{C}$ 上的约束是否成立，即可说明这一点。需要注意的是，当 $\mathcal{M}_Y$ 是一个退化族时，其上的所有不等式约束都会变成包含在简化行阶梯形中的等式约束，因此无需检查。于是，剩下需要检查的只有$\mathbf{Ac}$ 上的约束，列举如下

$$\lambda_X^{\min} \le c_0 + \mathbb{P}(Y=1)\{c_1+c_4+c_5\} + \mathbb{P}(Y=0)\{c_2+c_8+c_{10}\} \le \lambda_X^{\max} \tag{110}$$
$$1 - p_{00} - p_{01} + \lambda_X^{\min} \le \mathbb{P}(Y=0)\{c_5+c_7+c_{13}\} + \mathbb{P}(Y=1)\{c_{10}+c_{11}+c_{14}\} + c_{15} \le 1 - p_{00} - p_{01} + \lambda_X^{\max} \tag{111}$$
$$p_{00} - \lambda_X^{\max} \le \mathbb{P}(Y=0)\{c_4+c_6+c_{14}\} + \mathbb{P}(Y=1)\{c_8+c_9+c_{13}\} + c_{12} \le p_{00} - \lambda_X^{\min} \tag{112}$$
$$p_{01} - \lambda_X^{\max} \le \mathbb{P}(Y=0)\{c_1+c_9+c_{11}\} + \mathbb{P}(Y=1)\{c_2+c_6+c_7\} + c_3 \le p_{01} - \lambda_X^{\min} \tag{113}$$

Sin对于 $c_0 = c_1 = c_2 = c_3 = c_4 = c_5 = c_6 = c_7 = c_8 = c_9 = c_{12} = c_{13}$，其中每一项都可简化为

$$\lambda_X^{\min} \le \mathbb{P}(Y=0)c_{10} \le \lambda_X^{\max} \tag{114}$$
$$1 - p_{00} - p_{01} + \lambda_X^{\min} \le \mathbb{P}(Y=1)\{c_{10}+c_{11}+c_{14}\} + c_{15} \le 1 - p_{00} - p_{01} + \lambda_X^{\max} \tag{115}$$
$$p_{00} - \lambda_X^{\max} \le \mathbb{P}(Y=0)c_{14} \le p_{00} - \lambda_X^{\min} \tag{116}$$
$$p_{01} - \lambda_X^{\max} \le \mathbb{P}(Y=0)c_{11} \le p_{01} - \lambda_X^{\min} \tag{117}$$

根据定义，不等式 I 成立。根据 $\lambda_X^{\max} \ge \lambda_X^{\min}$的定义，不等式 III 和 IV 成立。对于不等式 II，结果如下:

$$\mathbb{P}(Y=1)\{c_{10}+c_{11}+c_{14}\} + c_{15} = 1 - p_{00} - p_{01} + \lambda_X^{\max} \tag{118}$$

这仍将满足这些不等式。这证明了我们针对这里所述情形的定理；对于其余情形，也可以重复类似的推导，即 $\mathbb{P}(Z=0|Y=0) = 0$, $\mathbb{P}(Z=0|Y=0) = 1$ 和$\mathbb{P}(Z=0|Y=1) = 1$。

17 定理 2 的证明

正如我们在引理 1 中所见，收紧 PNS 的界限等价于收紧 λ_X 的界限。通过求解以 $[\mathbf{Ac}]_0$ 为线性目标函数、并满足第 15 节中的多面体约束的线性规划，可以得到 λ_X 的界限。每个界限都是针对每一种可能的条件退化分布 $\mathbb{P}(Z,\ Y)$ 逐一推导得出的。这里，我们将 $\mathbb{P}(Z=0 \mid Y=1)$ 记为 p'_{01}，将 $\mathbb{P}(Z=0 \mid Y=0)$ 记为 p'_{00}，将 $\mathbb{P}(Z=0 \mid X=1)$ 记为 p_{01}，并将 $\mathbb{P}(Z=0 \mid X=0)$ 记为 p_{00}。这些界限通过单纯形算法推导得出，相关细节见 [41]。

17.1 $\mathbb{P}(Z=0 \mid Y=1)$

正如我们在第 16 节中所看到的，当 $\mathbb{P}(Z=0 \mid Y=1) = 0$ 时，关于 $\mathcal{C}$ 的等式约束可以更新为

$$c_{10} - c_{15} = \frac{p_{00} + p_{01} - \mathbb{P}(Y=0)}{\mathbb{P}(Y=0)} \tag{119}$$
$$c_{11} + c_{15} = \frac{\mathbb{P}(Y=0) - p_{00}}{\mathbb{P}(Y=0)} \tag{120}$$
$$c_{14} + c_{15} = \frac{\mathbb{P}(Y=0) - p_{01}}{\mathbb{P}(Y=0)} \tag{121}$$

此外，关于**Ac** 的以下不等式约束仍然成立，表示为

$$\lambda_X^{\min} \leq \mathbb{P}(Y=0)c_{10} \leq \lambda_X^{\max} \tag{122}$$

$$1-p_{00}-p_{01}+\lambda_X^{\min} \leq \mathbb{P}(Y=1)\{c_{10}+c_{11}+c_{14}\}+c_{15} \leq 1-p_{00}-p_{01}+\lambda_X^{\max} \tag{123}$$

$$p_{00}-\lambda_X^{\max} \leq \mathbb{P}(Y=0)c_{14} \leq p_{00}-\lambda_X^{\min} \tag{124}$$

$$p_{01}-\lambda_X^{\max} \leq \mathbb{P}(Y=0)c_{11} \leq p_{01}-\lambda_X^{\min} \tag{125}$$

为了应用单纯形算法，我们使用松弛变量和剩余变量，将最小化问题以标准形式表示为单纯形表，其中分别表示 s_i 和 e_i，其中 $s_i \geq 0 \quad \forall i$， e_i 也同理。

$$\begin{bmatrix}
z & c_{10} & c_{11} & c_{14} & c_{15} & s_1 & s_2 & s_3 & s_4 & e_1 & e_2 & e_3 & e_4 & RHS \\
1 & -\mathbb{P}(Y=0) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{p_{00}+p_{01}-\mathbb{P}(Y=0)}{\mathbb{P}(Y=0)} \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\mathbb{P}(Y=0)-p_{00}}{\mathbb{P}(Y=0)} \\
0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\mathbb{P}(Y=0)-p_{01}}{\mathbb{P}(Y=0)} \\
0 & \mathbb{P}(Y=0) & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_X^{\max} \\
0 & \mathbb{P}(Y=1) & \mathbb{P}(Y=1) & \mathbb{P}(Y=1) & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1-p_{00}-p_{01}+\lambda_X^{\max} \\
0 & 0 & 0 & \mathbb{P}(Y=0) & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & p_{00}-\lambda_X^{\min} \\
0 & 0 & \mathbb{P}(Y=0) & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & p_{01}-\lambda_X^{\min} \\
0 & \mathbb{P}(Y=0) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & \lambda_X^{\min} \\
0 & \mathbb{P}(Y=1) & \mathbb{P}(Y=1) & \mathbb{P}(Y=1) & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & \lambda_X^{\min}-p_{00}-p_{01}+1 \\
0 & 0 & 0 & \mathbb{P}(Y=0) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & p_{00}-\lambda_X^{\max} \\
0 & 0 & \mathbb{P}(Y=0) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & p_{01}-\lambda_X^{\max}
\end{bmatrix} \tag{126}$$

在指定基变量和非基变量时，我们必须谨慎选择正确的组合，以确保最终得到一个基本可行解。例如，c_{10}、 c_{11}、 c_{14} 和 c_{15} 不能全部选作非基变量，因为这将导致违反单纯形约束。

因此，我们知道必须选择松弛变量或剩余变量作为非基变量。根据定理**1**，我们知道，当该方程组有解时，$\lambda_X^{\max}$始终是该方程组的一个基本可行解。因此，我们继续将该方程组转换为规范形，以 s_1 和其他松弛变量、剩余变量的某种组合作为非基变量。相应的单纯形表就变为

$$\begin{bmatrix}
z & c_{10} & c_{11} & c_{14} & c_{15} & s_1 & s_2 & s_3 & s_4 & e_1 & e_2 & e_3 & e_4 & RHS \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_X^{\max} \\
0 & 1 & 0 & 0 & 0 & \frac{1}{\mathbb{P}(Y=0)} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\lambda_X^{\max}}{\mathbb{P}(Y=0)} \\
0 & 0 & 1 & 0 & 0 & -\frac{1}{\mathbb{P}(Y=0)} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{p_{01}-\lambda_{\max X}}{\mathbb{P}(Y=0)} \\
0 & 0 & 0 & 1 & 0 & -\frac{1}{\mathbb{P}(Y=0)} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{p_{00}-\lambda_X^{\max}}{\mathbb{P}(Y=0)} \\
0 & 0 & 0 & 0 & 1 & \frac{1}{\mathbb{P}(Y=0)} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\lambda_{\text{最大值}X} \quad +\mathbb{P}(Y=0)-p_{00}-p_{01}}{\mathbb{P}(Y=0)} \\
0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \lambda_X^{\max}-\lambda_X^{\min} \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \lambda_X^{\max}-\lambda_X^{\min} \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & \lambda_X^{\min}-\lambda_X^{\max} \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & \lambda_X^{\min}-\lambda_X^{\max} \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0
\end{bmatrix} \tag{127}$$

然而，该单纯形表中的松弛变量和剩余变量存在冗余，我们将对其进行系统性分析并予以剔除。首先， R_{12}, R_{11}, R_6 均为 $s_1 - s_2 = 0$ 的形式，不会限制 s_1 的取值范围，并且在单纯形算法的假设下始终可满足。

接下来， R_7， R_8， R_9， R_{10} 均表示同一个约束，即 $s_1 \leq \lambda_X^{\max} - \lambda_X^{\min}$，因此我们只保留 $R7$，得到如下化为规范形的简化单纯形表：

$$\begin{bmatrix} z & c_{10} & c_{11} & c_{14} & c_{15} & s_1 & s_3 & RHS \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & \lambda_X^{\max} \\ 0 & 1 & 0 & 0 & 0 & \frac{1}{\mathbb{P}(Y=0)} & 0 & \frac{\lambda_X^{\max}}{\mathbb{P}(Y=0)} \\ 0 & 0 & 1 & 0 & 0 & -\frac{1}{\mathbb{P}(Y=0)} & 0 & \frac{p_{01}-\lambda_X^{\max}}{\mathbb{P}(Y=0)} \\ 0 & 0 & 0 & 1 & 0 & -\frac{1}{\mathbb{P}(Y=0)} & 0 & \frac{p_{00}-\lambda_X^{\max}}{\mathbb{P}(Y=0)} \\ 0 & 0 & 0 & 0 & 1 & \frac{1}{\mathbb{P}(Y=0)} & 0 & \frac{\lambda_X^{\max}+\mathbb{P}(Y=0)-p_{00}-p_{01}}{\mathbb{P}(Y=0)} \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & \lambda_X^{\max} - \lambda_X^{\min} \end{bmatrix} \tag{128}$$

此处非基变量为 s_1，基变量为 z， c_{10}， c_{11}， c_{14}， c_{15}， s_2，基本可行解为

$$s_1 = 0 \tag{129}$$
$$z = \lambda_X^{\max} \tag{130}$$
$$c_{10} = \frac{\lambda_X^{\max}}{\mathbb{P}(Y=0)} \tag{131}$$
$$c_{11} = \frac{p_{01} - \lambda_X^{\max}}{\mathbb{P}(Y=0)} \tag{132}$$
$$c_{14} = \frac{p_{00} - \lambda_X^{\max}}{\mathbb{P}(Y=0)} \tag{133}$$
$$c_{15} = \frac{\lambda_X^{\max} + \mathbb{P}(Y=0) - p_{00} - p_{01}}{\mathbb{P}(Y=0)} \tag{134}$$
$$s_2 = \lambda_X^{\max} - \lambda_X^{\min} \tag{135}$$

可以验证，在定理1所列条件下，所有变量均为正。由于第1行中非基变量的系数并非非正数，因此这不是最小值。于是我们使用比值检验来确定将哪个变量设为非基变量。

• 对 $R2$进行比值检验：

$$c_{10} + \frac{s_1}{\mathbb{P}(Y=0)} = \frac{\lambda_X^{\max}}{\mathbb{P}(Y=0)} \tag{136}$$
$$c_{10} = \frac{\lambda_X^{\max} - s_1}{\mathbb{P}(Y=0)} \geq 0 \tag{137}$$
$$s_1 \leq \lambda_X^{\max} \tag{138}$$

• 对 $R3$进行比值检验

$$c_{11} - \frac{s_1}{\mathbb{P}(Y=0)} = \frac{p_{01} - \lambda_X^{\max}}{\mathbb{P}(Y=0)} \tag{139}$$
$$c_{11} = p_{01} - \lambda_X^{\max} + s_1 \geq 0 \tag{140}$$
$$s_1 \geq 0 \tag{141}$$

• 同样， $R4$ 也将施加 $s_1 \geq 0$

• 对 $R5$进行比值检验

$$c_{15} + \frac{s_1}{\mathbb{P}(Y=0)} = \frac{\lambda_X^{\max} + \mathbb{P}(Y=0) - p_{00} - p_{01}}{\mathbb{P}(Y=0)} \tag{142}$$
$$c_{15} = \frac{\lambda_X^{\max} + \mathbb{P}(Y=0) - p_{00} - p_{01} - s_1}{\mathbb{P}(Y=0)} \geq 0 \tag{143}$$
$$s_1 \leq \lambda_X^{\max} + \mathbb{P}(Y=0) - p_{00} - p_{01} \tag{144}$$

• 比值检验适用于 $R6$

$$s_1 + s_2 = \lambda_X^{\max} - \lambda_X^{\min} \geq 0 \tag{145}$$
$$s_1 \leq \lambda_X^{\max} - \lambda_X^{\min} \tag{146}$$

比值检验的胜出者，以及因此我们将其作为枢轴变量的变量，将由 $\min\{\lambda_X^{\max}, \lambda_X^{\max} + \mathbb{P}(Y=0) - p_{00} - p_{01}, \lambda_X^{\max} - \lambda_X^{\min}\}$ 决定。这里的最小值取决于 p_{00}、p_{01} 和 $\mathbb{P}(Y=0)$ 的取值，具体分为以下三种情况

- $p_{00}+p_{01}-1>0$：此处 $\lambda_{\text{最小值}_X} = p_{00}+p_{01}-1$。最小值将为 $\lambda_{\text{最大值}_X} + \mathbb{P}(Y=0) - p_{00} - p_{01}$，因此$c_{15}$ 将成为非基变量。
- $p_{00}+p_{01}-1=0$：此处 $\lambda_{\text{最小值}_X} = 0$。那么 c_{15} 也将成为非基变量。
- $p_{00}+p_{01}-1<0$：此处 λ最小值$_X = 0$。这归结为检查 λ最大值x与 λ最大值$_X + \mathbb{P}(Y=0) - p_{00} - p_{01}$ 之间的关系，具体情况如下文逐一分析。– λ最大值$x > \lambda$最大值$_X + \mathbb{P}(Y=0) - p_{00} - p_{01}$：这意味着 $p_{00}+p_{01}-1+\mathbb{P}(Y=1) > 0$，因此 c_{15} 将成为非基变量。– λ最大值$x = \lambda$最大值$_X + \mathbb{P}(Y=0) - p_{00} - p_{01}$：这意味着 $p_{00}+p_{01}-1+\mathbb{P}(Y=1) = 0$，此处 c_{10} 将成为非基变量。– λ最大值$x < \lambda$最大值$_X + \mathbb{P}(Y=0) - p_{00} - p_{01}$：这意味着 $p_{00}+p_{01}-1+\mathbb{P}(Y=1) < 0$，此处 c_{10} 将成为非基变量。

因此，视情况而定，我们将在 c_{10} 或 c_{15}处进行枢轴变换。以 c_{15} 为枢轴变换后的单纯形表为 a s

$$\begin{bmatrix} z & c_{10} & c_{11} & c_{14} & c_{15} & s_1 & s_3 & RHS \\ 1 & 0 & 0 & 0 & -\mathbb{P}(Y=0) & 0 & 0 & p_{00}+p_{01}-1+\mathbb{P}(Y=1) \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 & \frac{p_{00}+p_{01}-\mathbb{P}(Y=0)}{\mathbb{P}(Y=0)} \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & \frac{\mathbb{P}(Y=0)-p_{00}}{\mathbb{P}(Y=0)} \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & \frac{\mathbb{P}(Y=0)-p_{01}}{\mathbb{P}(Y=0)} \\ 0 & 0 & 0 & 0 & \mathbb{P}(Y=0) & 1 & 0 & \lambda_X^{\max} - p_{00} - p_{01} + \mathbb{P}(Y=0) \\ 0 & 0 & 0 & 0 & -\mathbb{P}(Y=0) & 0 & 1 & p_{00}+p_{01}-\mathbb{P}(Y=0)-\lambda_X^{\min} \end{bmatrix} \tag{147}$$

相应的基本可行解为

$$c_{15} = 0 \tag{148}$$
$$z = p_{00}+p_{01}-1+\mathbb{P}(Y=1) \tag{149}$$
$$c_{10} = \frac{p_{00}+p_{01}-\mathbb{P}(Y=0)}{\mathbb{P}(Y=0)} \tag{150}$$
$$c_{11} = \frac{\mathbb{P}(Y=0)-p_{00}}{\mathbb{P}(Y=0)} \tag{151}$$
$$c_{14} = \frac{\mathbb{P}(Y=0)-p_{01}}{\mathbb{P}(Y=0)} \tag{152}$$
$$s_3 = p_{00}+p_{01}-\mathbb{P}(Y=0)-\lambda_X^{\min} \tag{153}$$

所有数值均为正，因此这是一个基本可行解；由于非基变量的系数非正，所以这是一个最小值。相应的最小值为 $p_{00}+p_{01}-1+\mathbb{P}(Y=1)$。
同样，当 c_{10} 被换入基时，对应的单纯形表为

$$\begin{bmatrix} z & c_{10} & c_{11} & c_{14} & c_{15} & s_1 & s_3 & RHS \\ 1 & -\mathbb{P}(Y=0) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbb{P}(Y=0) & 0 & 0 & 0 & 1 & 0 & \lambda_X^{\max} \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & \frac{p_{01}}{\mathbb{P}(Y=0)} \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & \frac{p_{00}}{\mathbb{P}(Y=0)} \\ 0 & -1 & 0 & 0 & 1 & 0 & 0 & \frac{\mathbb{P}(Y=0)-p_{00}-p_{01}}{\mathbb{P}(Y=0)} \\ 0 & -\mathbb{P}(Y=0) & 0 & 0 & 0 & 0 & 1 & -\lambda_X^{\min} \end{bmatrix} \tag{154}$$

在这种情况下，基本可行解为

$$z = 0 \tag{155}$$
$$c_{10} = 0 \tag{156}$$
$$s_1 = \lambda_X^{\max} \tag{157}$$
$$c_{11} = \frac{p_{01}}{\mathbb{P}(Y=0)} \tag{158}$$
$$c_{14} = \frac{p_{00}}{\mathbb{P}(Y=0)} \tag{159}$$
$$c_{15} = \frac{\mathbb{P}(Y=0) - p_{00} - p_{01}}{\mathbb{P}(Y=0)} \tag{160}$$
$$s_2 = -\lambda_X^{\min} = 0 \tag{161}$$

这两个条件概括了范围限制 λ_X。综合起来， λ_X 的范围限制取决于 $\Phi_X = p_{00} + p_{01} - 1$ 的值，具体如下：

- $\Phi_X \geq 0$： λ_X 的更新范围为 $[\Phi_X + \mathbb{P}（Y=1）\ \lambda_{最大值X}\ ]$。
- $\Phi_X < 0$：这分为两种情况：– 如果 $\Phi_X + \mathbb{P}（Y=1）\leq 0$，那么 λ_X 的更新范围为 $[\lambda_{最小值X}，\ \lambda_{最大值X}\ ]$。– $\Phi_X + \mathbb{P}（Y=1）> 0$： 的最大值为 $\lambda_{最大值X}$ 等于 $[\Phi_X + \mathbb{P}（Y=1）\ \lambda_{最大值X}\ ]$

类似的推导可得出 λ_X 在 $\mathbb{P}（Z，\ Y）$具有退化条件分布时，各种情况下范围限制的等价结果。推导过程留作读者练习，下面仅给出结果。

17.2 关于 $p'_{00} = 0$的定理陈述

当 $\lambda_Y^{\max} = \lambda_Y^{\min} = p'_{00} = 0$ 且满足分布兼容性中概述的条件时，对 λ_X 取值范围施加反事实一致性的要求，取决于归一化效应强度 Φ_X 的值，具体如下：

- $\Phi_X \geq 0$：更新后的范围将为 λ_X 为 $[\Phi_X + \mathbb{P}(Y=0)\ \lambda_{最大值X}\ ]$
- $\Phi_X < 0$：分为以下两种情况：– $\Phi_X + \mathbb{P}(Y=0) \leq 0$：此时不会发生证伪， λ_X 的范围将为 $[0, \lambda_{最大值X}\ ]$。– $\Phi_X + \mathbb{P}(Y=0) > 0$： λ_X 的更新后范围将为 $[\Phi_X + \mathbb{P}(Y=0)\ \lambda_{最大值X}\ ]$。

17.3 关于 $p'_{00} = 1$的定理陈述

结果分为以下几种情况

- $\Phi_X \leq 0$：然后， λ_X 的更新范围为 $[\mathbb{P}(Y=0)\ \lambda_{最大值X}\ ]$。
- $\Phi_X > 0$：这分为两种情况– $\Phi_X - \mathbb{P}(Y=0) \geq 0$：则不会发生证伪，且 λ_X 的范围为 $[\Phi_X, \lambda_{最大值X}\ ]$。– $\Phi_X - \mathbb{P}(Y=0) < 0$：的更新范围将为 $[\mathbb{P}(Y=0)\ \lambda_{最大值X}\ ]$

17.4 关于 $p'_{01} = 1$的定理陈述

结果同样分为以下几种情况

- $\Phi_X \leq 0$：然后， λ_X 的更新范围为 $[\mathbb{P}(Y=1)\ \lambda_{最大值X}\ ]$。
- $\Phi_X > 0$：分为两种情况– $\Phi_X - \mathbb{P}(Y=1) \geq 0$：此时没有证伪，且 λ_X 的取值范围为 $[\Phi_X, \lambda_{\max X}\ ]$。– $\Phi_X - \mathbb{P}(Y=1) < 0$：更新后的 λ_X 取值范围为 $[\mathbb{P}(Y=1)\ \lambda_{\max X}\ ]$

17.5 界限的重述 λ_X

当 $\mathbb{P}(Z,\ Y)$ 具有退化条件分布时，上述四个结果都可以用指示函数简洁地表述如下：更新后的 λ_X 下界可以表述为

$$\mathbb{D}_0 \equiv \mathbb{I}(p'_{00}=0)\mathbb{P}(Y=0)+\mathbb{I}(p'_{01}=0)\mathbb{P}(Y=1) \tag{162}$$
$$\mathbb{D}_1 \equiv \mathbb{I}(p'_{00}=1)\mathbb{P}(Y=0)+\mathbb{I}(p'_{01}=1)\mathbb{P})(Y=1) \tag{163}$$
$$\max\{\Phi_X+\mathbb{D}_0,\mathbb{D}_1\} \leq \lambda_X \tag{164}$$

17.6 PNS 的更新界限

由于 PNS 给定为 $p_{00}-\lambda_X$，这些界限可以直接转换。由于 λ_X 的上界保持不变，PNS 的下界也保持不变。不过，PNS 的上界会降低，下面给出更新后的 PNS 界限。

$$\max[0,p_{11}-p_{10}] \leq PNS \leq \min\{p_{00}-\mathbb{D}_1,p_{11}-\mathbb{D}_0\} \tag{165}$$

18 扩展至其他反事实概率

虽然我们将讨论重点放在 PNS 上，但我们对某些 λ_X 取值所对应的结构因果模型进行的证伪，也可用于界定其他反事实量，下面将讨论一个例子。我们将通过另外两个值得关注的反事实量来说明这一点，它们的灵感分别来自 PN 和 PS，并采用类似的命名方式

• 充分非单调性概率

$$\mathbb{P}(Z=0 \mid X=0,Z=1,do(X=1)) \tag{166}$$

• 必要非单调性概率

$$\mathbb{P}(Z=1 \mid X=1,Z=0,do(X=0)) \tag{167}$$

18.0.1 充分非单调性概率

这可以通过 Pearl 的溯因、干预和预测三步过程来计算 [1]。我们首先使用贝叶斯定理计算$\mathbb{P}(N_Z \mid X=0,Z=1)$

$$\mathbb{P}(N_Z \mid X=0,Z=1) = \frac{\mathbb{P}(X=0,Z=1 \mid N_Z)\mathbb{P}(N_Z)}{\mathbb{P}(X=0,Z=1)} \tag{168}$$
$$= \frac{\mathbb{P}(Z=1 \mid X=0,N_Z)\mathbb{P}(X=0 \mid N_Z)\mathbb{P}(N_Z)}{\mathbb{P}(X=0,Z=1)} \tag{169}$$
$$= \frac{\mathbb{P}(Z=1 \mid X=0,N_Z)\mathbb{P}(N_Z)}{\mathbb{P}(Z=1 \mid X=0)} \tag{170}$$
$$\tag{171}$$

因此 $\mathbb{P}(N_Z \mid X=0,Z=1) = [0,\ \frac{1-p_{00}-p_{01}-\lambda_X}{\mathbb{P}(Z=1|X=0)}\ 0,\ \frac{p_{01}-\lambda_X}{(Z=1|X=0)}]$。由此，反事实概率的值可表示为

$$\mathbb{P}(Z=0 \mid X=0,Z=1,do(X=1)) = \frac{p_{01}-\lambda_X}{\mathbb{P}(Z=1 \mid X=0)} \tag{172}$$

18.0.2 必要非单调性概率

按照类似的步骤，我们得到

$$\mathbb{P}(Z=1 \mid X=1,Z=0,do(X=0)) = \frac{p_{01}-\lambda_X}{\mathbb{P}(Z=0 \mid X=1)} \tag{173}$$

因此我们看到，通过限制 λ_X 的取值范围，也可以对其他反事实设定界。

参考文献

[1] Judea Pearl。《因果性》。剑桥大学出版社，2009年。[2] Scott Mueller 和 Judea Pearl。《个性化决策：概念性介绍》。arXiv 预印本 arXiv:2208.09558，2022年。[3] David L Faigman、John Monahan 和 Christopher Slobogin。《科学专家证词中的群体到个体（g2i）推断》。《芝加哥大学法律评论》，第417–480页，2014年。[4] Judea Pearl。《因果概率：三种反事实解释及其识别》。载于《概率与因果推断：Judea Pearl 的著作》，第317–372页。2022年。[5] Muin J Khoury、W Dana Flanders、Sander Greenland 和 Myron J Adams。《流行病学研究中易感性的测量》。《美国流行病学杂志》，129(1):183–190，1989年。[6] Charles F Manski。《治疗效应的非参数界限》。《美国经济评论》，80(2):319–323，1990年。[7] Jin Tian 和 Judea Pearl。《因果概率：界限与识别》。《数学与人工智能年刊》，28(1):287–313，2000年。[8] Alexander Balke 和 Judea 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。[9] James Robins 和 Sander Greenland。《个体风险随机模型下的因果概率》。《生物统计学》，第1125–1138页，1989年。[10] A Philip Dawid、Monica Musio 和 Rossella Murtas。《因果概率》。《法律、概率与风险》，16(4):163–179，2017年。[11] Junzhe Zhang、Jin Tian 和 Elias Bareinboim。《基于观察数据和实验数据的部分反事实识别》。载于《国际机器学习会议》，第26548–26558页。PMLR，2022年。[12] Maria Cuellar。《法律中的因果推理与数据分析：因果概率的定义、估计与应用》。因果概率的估计与应用（2018年5月31日），2018年。[13] Kirtan Padh、Jakob Zeitler、David Watson、Matt Kusner、Ricardo Silva 和 Niki Kilbertus。《用于界定治疗效应的随机因果规划》。arXiv 预印本 arXiv:2202.10806，2022年。[14] Guilherme Duarte、Noam Finkelstein、Dean Knox、Jonathan Mummolo 和 Ilya Shpitser。《离散情境下因果推断的自动化方法》。arXiv 预印本 arXiv:2109.13471，2021年。[15] Michael C Sachs、Gustav Jonzon、Arvid Sj¨olander 和 Erin E Gabriel。《推导因果效应紧确符号界的一般方法》。《计算与图形统计学杂志》，（即将发表）：1–23，2022年。[16] Ang Li 和 Judea Pearl。《非二元治疗和效应的因果概率》。arXiv 预印本 arXiv:2208.09568，2022年。[17] Luigi Gresele、Julius Von K¨ugelgen、Jonas K¨ubler、Elke Kirschbaum、Bernhard Sch¨olkopf 和 Dominik Janz-ing。《通过结构因果边缘问题进行因果推断》。载于《国际机器学习会议》，第7793–7824页。PMLR，2022年。[18] Antti Hyttinen、Frederick Eberhardt 和 Patrik O Hoyer。《利用潜变量学习线性循环因果模型》。《机器学习研究杂志》，13(1):3387–3439，2012年。[19] Shohei Shimizu、Takanori Inazumi、Yasuhiro Sogawa、Aapo Hyv¨arinen、Yoshinobu Kawahara、Takashi Washio、Patrik O Hoyer 和 Kenneth Bollen。《DirectLiNGAM：学习线性非高斯结构方程模型的直接方法》。《机器学习研究杂志》，12:1225–1248，2011年。[20] ET Jaynes。《贝叶斯方法与最大熵方法的关系》。载于《科学与工程中的最大熵和贝叶斯方法》，第25–29页。施普林格，1988年。[21] Dominik Janzing。《最大熵的因果版本与理由不足原则》。《因果推断杂志》，9(1):285–301，2021年。[22] Sergio H Garrido Mejia、Elke Kirschbaum 和 Dominik Janzing。《通过使用最大熵合并数据集获取因果信息》。载于《国际人工智能与统计会议》，第581–603页。PMLR，2022年。[23] Jakob Zeitler 和 Ricardo Silva。《用于界定治疗效应的因果边缘多面体》。arXiv 预印本 arXiv:2202.13851，2022年。

[24] R. Chaves、L. Luft、TO. Maciel、D. Gross、D. Janzing 和 B. Sch¨olkopf。通过信息不等式推断潜在结构。载于 NL Zhang 和 J Tian 编，《第三十届人工智能不确定性会议论文集》，第112–121页，俄勒冈州科瓦利斯，2014年。AUAI出版社。[25]James Robins。针对持续暴露期的死亡率研究中因果推断的新方法——应用于控制健康工人幸存者效应。《数学建模》，7(9-12):1393–1512，1986年。[26] Peter Spirtes、Clark N Glymour、Richard Scheines 和 David Heckerman。《因果关系、预测与搜索》。麻省理工学院出版社，2000年。[27] Jonas Peters、Dominik Janzing 和 Bernhard 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。[28] T. Cover 和 J. Thomas。《信息论要素》。Wileys通信系列，纽约，1991年。[29] D. Janzing、D. Balduzzi、M. Grosse-Wentrup 和 B. Sch¨olkopf。量化因果影响。《统计学年鉴》，41(5):2324–2358，2013年。[30] Edwin T Jaynes。《概率论：科学逻辑》。剑桥大学出版社，2003年。[31] John Shore 和 Rodney Johnson。最大熵原理与最小交叉熵原理的公理化推导。《IEEE信息论汇刊》，26(1):26–37，1980年。[32] Jos Uffink。最大熵原理能否被解释为一致性要求？《科学史与科学哲学研究B辑：现代物理学科学史与科学哲学研究》，26(3):223–261，1995年。[33] Peter D Gr¨unwald 和 A Philip Dawid。博弈论、最大熵、最小差异与稳健贝叶斯决策理论。《统计学年鉴》，32(4):1367–1433，2004年。[34] Xingsi Li 和 Shaohua Pan。熵优化的对偶问题及其应用。载于《非线性力学中的互补性、对偶性与对称性》，第167–178页。施普林格，2004年。[35] Elias Bareinboim 和 Judea Pearl。因果推断与数据融合问题。《美国国家科学院院刊》，113(27):7345–7352，2016年。[36] Paul H¨unermund 和 Elias Bareinboim。计量经济学中的因果推断与数据融合。arXiv预印本 arXiv:1912.09104，2019年。[37] Peter C Taylor、Edward C Keystone、Desiree Van Der Heijde、Michael E Weinblatt、Liliana del Carmen Morales、Jaime Reyes Gonzaga、Sergey Yakushin、Taeko Ishii、Kahaku Emoto、Scott Beattie 等。巴瑞替尼与安慰剂或阿达木单抗治疗类风湿关节炎的比较。《新英格兰医学杂志》，376(7):652–662，2017年。[38] Mark C Genovese、Kenneth Kalunian、Jacques-Eric Gottenberg、Neelufar Mozaffarian、Beatrix Bartok、Franziska Matzkies、Jie Gao、Ying Guo、Chantal Tasset、John S Sundy 等。非戈替尼与安慰剂对改善病情抗风湿药治疗无效的中重度类风湿关节炎患者临床应答的影响：FINCH 2随机临床试验。《美国医学会杂志》，322(4):315–325，2019年。[39] David T Felson、Jennifer J Anderson、Maarten Boers、Claire Bombardier、Daniel Furst、Charles Goldsmith、Linda M Katz、Robert Lightfoot Jr、Harold Paulus、Vibeke Strand 等。美国风湿病学会对类风湿关节炎改善情况的初步定义。《关节炎与风湿病：美国风湿病学会官方期刊》，38(6):727–735，1995年。[40] Fernando P Polack、Stephen J Thomas、Nicholas Kitchin、Judith Absalon、Alejandra Gurtman、Stephen Lockhart、John L Perez、Gonzalo P´erez Marc、Edson D Moreira、Cristiano Zerbini 等。BNT162b2信使RNA新冠疫苗的安全性和有效性。《新英格兰医学杂志》，2020年。[41] George B Dantzig。单纯形法的起源。载于《科学计算史》，第141–151页，1990年。