# 因果确证测度:从辛普森悖论到 COVID-19

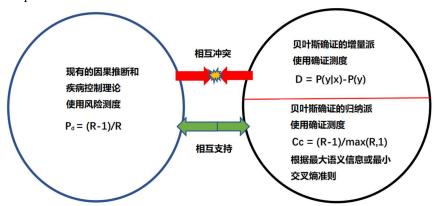
英文原文 Entropy | Free Full-Text | Causal Confirmation Measures: From Simpson's Paradox to COVID-19 (mdpi.com)

#### 鲁晨光

**摘要:**用统计数据比较两种原因对结果的影响时,如果我们得到的合并结论和分组结论相反,这被称之为辛普森悖论(英文是 Simpson Paradox,缩写为 SP)。现在流行的因果推断理论(英文是 Popular Causal Inference Theory,缩写为 PCIT)通过去混杂影响,使合并结论和分组结论一致,从而消除 SP。PCIT 使用相对风险差  $P_a$ =max(0,(R-1)/R)作为反映因果关系的概率(R=正例比例/反例比例),而哲学家 Fitelson 用确证测度 D (D=后验概率-先验概率)评价因果关系强度。Fitelson 得出结论:从贝叶斯确证理论看,我们无需考虑混杂影响,直接接受合并结论就行了。为了消除 PCIT 和贝叶斯确证之间的矛盾,作者使用语义信息方法推导出因果确证测度 Cc=R-1)/max(R,1)。这一测度很像  $P_a$ ,但是具有归一化性质(在-1和1之间变化)和原因对称性(相反原因会使得确证度为负)。Cc 特别适合原因抑制结果(比如疫苗抑制感染)的情况(确证度是负的)。文中提供的一些例子(关于肾结石治疗和 COVID-19)表明  $P_a$ 和 Cc 比 D 更加合理,Cc 比  $P_a$  更加有用。 $P_a$ 和 Cc 的合理性反过来支持了贝叶斯确证的归纳派。

**关键词:** 因果确证; 贝叶斯确证; 因果推理,语义信息测度,交叉熵,辛普森悖论,风险测度

### Graphical Abstract:



## 1. 引言

因果确证既是贝叶斯确证的扩充,也是因果推断的任务之一。虽然现存的因果推断理论(英文是 Existing Causal Inference Theory,后面缩写为 ETCI)——包含 Rubin 的潜在结果模型【1,2】和 Pearl 的因果结构模型(即因果图)【3,4】——已经取得很大成功,但是它很少提及因果确证。

贝叶斯确证理论也被称之为确证理论,可分为增量派和归纳派。增量派肯定确证是度量证据 e 对假设 h 支持的强度,如 Fitelson 所述【5】。增量派跟随卡尔纳普【6】用假设的概率(或逻辑概率)的增量  $P(h \mid e)$ -P(h)作为确证测度。增量派代 1/21

表 Fitelson 使用该测度讨论因果确证,得出和 ECIT 不兼容的结论【5】。另一方面,归纳派【7,8】把确证当做归纳的现代形式,其任务是度量大前提的可信度——样本或样本分布支持的可信度。

我们用 e->h 表示一个大前提,变量 e 的两个可能取值是 e1 和它的否定 e0, 变量 h 的两个可能取值是 h1 和它的否定 h0. 则一个样本由四个样例(e1, h1)、(e1, h0)、(e0, h1)和(e0, h0)按不同比例构成。确证度就是样本支持的大前提的可信度,我们可以利用确证度做模糊三段论推理即概率预测【8】。归纳派的研究者通常使用正例和反例的比例(P(e1 h1)和 P(e1 h0))或似然比(R=P(e1 h1)/P(e1 h0))表达确证测度。

一个确证测度通常被表示为 C(e, h)或 C(h, e)。笔者持归纳派观点,为使任务清晰而建议使用  $C(e\rightarrow h)$ 表示确证度【8】。本文还使用 x=>y 表示"原因 x 导致结果 y。"

虽然两派对确证任务理解不同,但是都利用样本分布 P(e, h)构造确证测度。两派已经提供很多确证测度【8,9】。大多数人都赞成确证测度应有两个好性质:

- 归一化性质【9,10】:即 C(e, h)或 C(e-h)在-1 和 1 之间变化,能反映一个推理规则 e-h 和最可信和最不可信规则之间的差距)。
- 假设对称性【11】或后件对称性【8】,即  $c(e_1->h_1)=-c(e_1->h_0)$ 。比如: C(乌鸦->-是黑的)=-C(乌鸦->不是黑的)。

笔者曾提出具有对称性和归一化性质的两个确证测度:信道确证测度  $b^*(e - h)$  和预测确证测度  $c^*(e - h)$  【8】,分别反映 e - h 作为信道的可信性和作为预测的可信性。两者都有上述好性质,并且可以用于概率预测——根据证据 e 预测假设 h 发生的概率。

贝叶斯确证要确证的是关联关系——不同于因果关系。关联关系包含因果关系,但是很多关联关系并不是因果关系。一个原因是:关联关系的存在是对称的,比如有  $P(h|e)\neq0$  就有  $P(e|h)\neq0$ ,而因果关系的存在是不对称的。 比如:对于医学中, P(检验阳性|感染)即表示关联关系,也表示因果关系。但是反过来, <math>P(e) 染 | 阳性)只表示关联关系,而不表示因果关系。 另一个例子是:关联的两方 A 和 B(比如:空调好卖和电扇好卖)的原因是第三方 C(天热): A 和 B 都是结果, P(B|A)和 P(A|B)都不表示因果关系。

因果推断只考虑自然和社会中的不确定因果关系,不考虑数学中的因果关系,比如(x+1)(x-1)<x²是因为(x+1)(x-1)=x²-1。 我们知道康德区分分析判断和综合判断。虽然因果推断是一个数学方法,但是它用于综合判断——为了获得生物学,心理学,经济学等领域的不确定规律。因果确证只考虑二元因果关系。

虽然 ECIT 很少谈及"因果确证",但是流行病学和因果推断研究者已经使用了一些有确证含义的测度,包括风险差【12】:

$$RD = P(y_1 \mid x_1) - P(y_1 \mid x_0), \tag{1}$$

相对风险差或风险比率(类似于医学检验中似然比):

$$RR = P(y_1 \mid x_1) / P(y_1 \mid x_0), \tag{2}$$

和因果概率:  $P_d$  (见 Rubin 和 Greenland 文章【13】), 或原因必要性概率: PN (见 Pearl 【3】). 我们有:

$$P_d = PN = \max(0, \frac{P(y_1 \mid x_1) - P(y_1 \mid x_0)}{P(y_1 \mid x_1)}).$$
(3)

 $P_d$ 也叫相对风险差(英文是 Relative Risk Reduction,缩写为 RRR)。上面公式中

函数 Max(0, .)的最小值是 0,这是为了使  $P_d$  更像是概率。 $P_d$  和笔者提出的信道确证测度  $b^*$ 很相似, 不过  $b^*$ 在-1 和 1 之间变化。上面公式不仅能度量原因带来的风险或相对风险,也能度量原因带来的成功或相对成功。

上述测度很有实践意义,但是它们并不具有对称性和归一化性,不能说是很好的确证测度。如果仅仅用统计概率得到上面测度还存在辛普森悖论(Simpson's Paradox,缩写为SP)。就像乌鸦悖论一致伴随贝叶斯确证研究,SP一直伴随因果推断研究。辛普森【14】提出这一悖论时使用下面例子:

**例 1【15】**. 1951 年,加里福尼亚大学的贝克莱研究生院秋季招生,数据表明,44%的男性被录取,而只有 35%的女性被录取。有人质疑:这里可能存在性别歧视。然而,大多数系的女性申请人的录取率高于男性。

贝克莱大学是否存在性别偏见?我们应该接受分组结论(来自每个系)还是合并结论?根据合并结论,是存在对女性的偏见;但是根据分组结论则相反——学校对女性有所照顾。所以,我们说这里存在悖论。

这个例子有些复杂,争议较多。为了简化问题,我们使用因果推断研究者经常提到的肾结石手术例子——例 2——定量说明这一悖论。

我们用  $x_1$  表示唯一原因(或干预),用  $x_0$  表示缺省原因或无原因。如果比较两种原因的结果,我们可以用  $x_1$  和  $x_2$  或  $x_i$  和  $x_j$  表示,在这种情况下,我们可以假设其中一个是缺省原因,如  $x_0$ 。

**例 2 【16,17】**. 假设两种肾结石手术  $x_1$  和  $x_2$ 。我们根据病人结石尺寸把它们分成两组  $g_1$ (较小尺寸)和  $g_2$ (较大尺寸),  $y_1$  表示成功。则有可能出现图 1 所示情况:每一组都是  $x_2$  成功率高于  $x_1$  的成功率,但是合并结论却相反。

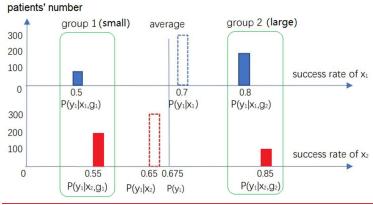


图 1. 图解辛普森悖论. 每一组中  $x_2$  的成功率  $P(y_1|x_2, g)$ 高于  $x_1$  的成功率( $P(y_1|x_1, g)$ ; 但是使用求重心方法得到  $x_2$  的总的成功率  $P(y_1|x_2)$ =0.65 却低于  $x_1$  的成功率  $P(y_1|x_1)$ =0.70.

根据 Rubin(或 Neyman 和 Rubin)的潜在结果模型【1】,我们应该选择分组结论,理由是结石大小是混杂因素,合并结论受了混杂因素影响。我们应该消除这种影响。消除的方法是想象每一组病人人数不变——不管我们使用  $x_1$  还是  $x_2$ . 然后我们用  $P(y_1)$ 取代  $P(y_i|x_1)$ 和  $P(y_i|x_2)$  (i=1, 2),求出两组新的平均成功率——Rubin 用  $P(y_1^{1})$ 和  $P(y_1^$ 

但是,是不是只要两种结论不一致就一定接受分组结论?未必!下面是一个 反例。 **例** 3 (来自【18】). 吃降压药记为  $x_1$ ,不吃降压药记为  $x_0$ 。 $y_1$ 表示恢复健康;  $y_0$  表示  $y_1$  的否定。病人被划分我高压人群( $g_1$ )和低压人群( $g_0$ ). 可能每一组都是  $P(y_1|g,x_1) < P(y_1|g,x_0)$ ,即不吃药好过吃药(因为吃药有副作用)。但是合并结论却是:  $P(y_1|x_1) > P(y_1|x_0)$ ,意味着吃药好过不吃药。

ECIT 告诉我们: 我们应该接受整体结论:  $x_1$  好过  $x_0$ , 因为血压是中间因素而不是混杂因素,也受吃药影响。 $x_1$  能把病人从高血压变为低血压,这是所期望的。我们不需要调整加权系数——用 P(g)代替  $P(g \mid x)$ 。两组病人数量在吃药后发生变化是正常的,我们应该接受合并结论而不是分组结论。分组结论  $P(y_1 \mid g, x_1)$  <  $P(y_1 \mid g, x_0)$ 反映的是: 降压药有副作用。

另外,也有些例子,从一个角度看分组结论是对的;而从另一个角度,合并结论是对的。

例 4 (来自【19】). 美国 2020 年 6 月的统计数据显示,COVID19 导致非西班牙裔白人(简称 NH 白人)病死率(Case Fatality Rate,英文缩写是 CFR)高于其他人(合并结论)。因为感染的人中,35.3%是非西班牙裔白人,而因 COVID19 死亡的人中,49.5 是 NH 白人。好像 COVID19 对白人来说更危险。但是 Dana Mackenzie 指出【19】:从各个年龄段看,结论相反——NH 白人致死率低于其他人。这正是 SP 的一个例子。原因是白人寿命较长,老年人相对比例比较大,而 COVID-19 对老年人来说更危险。

Kügelgen 等人【20】通过分析中国和意大利 2020 年报告的 COVID-19 病死率也指出这里存在 SP。虽然意大利总体病死率更高,但是从各个年龄组看,中国病死率更高。原因是感染者中,意大利老年人比例较大。

根据 Rubin 的潜在结果模型和 Pearl 的因果结构模型,如果我们认为 NH 白人寿命长的原因是医疗条件或其他因素而不是种族,人的寿命就是混杂因素。这时我们应该接受分组结论。另一方面,如果我们认为白人寿命长是因为他们是白人,那么寿命就是中介因素,这时我们应该接受整体结论。

例 1 类似于例 4,但是更加不容易理解。数据显示女性申请人有选择录取率低的专业的倾向(或许是因为较低的门槛导致更激烈竞争).这一倾向就像是白人的寿命,如果我们把它看做混杂因素,则贝克莱大学录取学生不但不存在对女性的偏见,而是有所照顾。另一方面,如果把女性看作这种倾向性的原因,则合并结论是对的,对女性的偏见是存在的。两种解释哪个是对的,取决于我们的视角。

Pearl 的因果图【3】能让我们更清楚看清:同样的数据,假设因果关系不同,结论也会不同;光有数据是不够的,我们还需要因果结构模型。

然而,增量派哲学家 Fitelson 争论说【5】: Pearl 的方法或许合适因果推断,但是从贝叶斯确证的角度看,我们不需要考虑因果关系,仅仅根据数据,接受合并结论就行了。从贝叶斯确证的角度看,SP 不存在。其理由是,我们可以用确证测度

measure 
$$i = P(y_1 | x_1) - P(y_1)$$
 (4)

(即【5】中的 *measure i*)度量因果关系,并且证明(见【5】中附录 Fact 3 的证明): 对于每一组,如果

$$P(y_1|x_1,g_i)>P(y_1|x_2,g_i), i=1,2,$$
 (5)

那么,必然有  $P(y_1|x_1)>P(y_1)$ . 上面">"换成"<",结论同样。所以,不像 RD 和  $P_d$ ,

D 不会引起悖论。

然而,上面不等式(5)是非常严格的条件,它已经排除所有产生悖论的联合概率分布 P(y, x, g),包括他自己简化的关于贝克莱大学男女录取率的联合概率分布。

我们不禁要问:

- 对于例 2, 我们也要接受合并结论,不考虑手术难度?这样合理吗?
- 这样的和因果推断理论及医学实践不兼容的确证测度有必要应用到因果确证吗?
- 除了不兼容的确证测度,难道就没有兼容的确证测度了? 除了增量派确证测度,还有归纳派的确证测度,比如 Kemeny 和 Oppenheim 于 1952 [12]提出的确证测度 F 和本文作者于 2020 年提出的确证测度 *b*\*。 本文主要目的是:
  - 结合 ETCI 和语义信息方法(使用最小交叉熵准则)推导出因果确证测度  $Cc(x_1=>y_1)$  ("C" 意思是确证,"c"指原因),它和  $P_d$ 类似,但是还能度量 负的因果关系(比如:疫苗=>感染);
  - 通过几个含有 SP 的例子说明测度 Cc 和  $P_a$ 比 measure i (即后面的 D) 更 合适于因果确证;
  - 支持贝叶斯确证的归纳派。

作者在提供  $b^*$ (信道确证测度)的同时提供了  $c^*$ (预测确证测度,用以消除乌鸦悖论)【8】. 推广  $c^*$ 到因果确证领域就得到确证测度 Ce--它反映结果的不可避免性或原因的充分性.

# 2. 背景

## 2.1. 贝叶斯确证: 增量派和归纳派

一个全称判断假设等价于一个假言判断或一个推理规则,比如"所有乌鸦是黑的"等价于"如果 x 是乌鸦,则 x 是黑的",我们简记为"乌鸦->黑的"。由于休谟和波普尔的批评,人们不再指望归纳出绝对正确的全称判断或大前提,而是希望得到它们的可信度,样本或样本分布支持的可信度就是确证度。

注意: 命题是不需要确证的, 比如"Tom 是老人",根据语言定义或用法,检查 Tom 的年龄就知道了。分析判断,比如"18 以上的人是成年人",也不需要确证, 根据语义或政府定义就知道了。只有综合判断(比如,"All ravens 是黑的","如果一个人核酸检验呈阳性,则他很可能被 COVID-19 被感染")需要确证。

一个自然的想法是用条件概率 P(h|e)确证推理规则  $e\rightarrow h$  或  $x\rightarrow y$ 。这一确证测度被 Fitelson 推荐,记为 f(e,h)(即 Fitelson 文中的 confirm f),即

f(e,h)=P(h|e). (Carnap, 1962 [6], Fitelson, 2017 [5])

但是 P(h|e)非常依赖 h 的先验概率 P(h). 比如:对于 COVID-19 流行的地方,P(h)就大,P(h|e)也大;而不流行的地方 P(h|e)就较小。所以 P(h|e)不能反映 e 的必要性。一个极端的例子是:h 和 e 完全无关,但是如果 P(h)很大,因为 P(h|e)=P(h,e)/P(e)=P(h),所以 P(h|e)也很大。这时候 P(h|e)并不反映"e->h"的可信度。比如:h="明天没有地震",其概率很大,假设是 P(h)=0.999;e="葡萄熟

了"。两者无关,但是 P(h|e)=P(h)=0.999. 这个值虽然很大, 但是不能说通过葡萄熟了可以预测有误地震。

为此,研究者们采用两个概率或条件概率的差表示可信度。至于是统计概率还是逻辑概率,他们没有严格区分。这些测度包括:

 $D(e_1, h_1) = P(h_1 | e_1) - P(h_1)$  (Carnap, 1962 [6]),  $M(e_1, h_1) = P(e_1 | h_1) - P(e_1)$  (Mortimer, 1988 [21]),  $R(e_1, h_1) = \log[P(h_1 | e_1)/P(h_1)]$  (Horwich, 1982 [22]),  $C(e_1, h_1) = P(h_1, e_1) - P(e_1) P(h_1)$  (Carnap,1962 [6]),

$$Z(h_1, e_1) = \begin{cases} [P(h_1 | e_1) - P(h_1)] / P(h_0, \text{ as } P(h_1 | e_1) \ge P(h_1), \\ [P(h_1 | e_1) - P(h_1)] / P(h_1), \text{ otherwise.} \end{cases}$$
 (Crupi *et al.*, 2007 [10]),

上面测度中  $D(e_1,h_1)$ 就是 Fitelson 建议的增量确证测度 cofirm i.  $R(e_1,h_1)$ 其实是一个信息测度,它可以写成  $\log P(h_1|e_1)-\log P(h_1)$ ,而  $\log P(h_1|e_1)-\log P(h_1)=\log P(e_1|h_1)-\log P(e_1)=\log P(h_1,e_1)-\log [P(h_1)P(e_1)]$ ,可见 D,M,and C 都和 R 正相关,可以相互取代。Z 则是 D 的归一化测度,保证确证度在-1 和 1 之间变化,且具有对称性【10】。因此,笔者称上述述确证度的支持者为增量派。

另一方面,归纳派两个条件概率(反映正例比例和反例比例)的差或似然比表示确证测度。这些测度包括:

 $S(e_1, h_1) = P(h_1 | e_1) - P(h_1 | e_0)$  (Christensen, 1999 [23]),

 $N(e_1, h_1) = P(e_1 | h_1) - P(e_1 | h_0)$  (Nozik, 1981 [24]),

 $L(e_1, h_1) = \log[P(e_1 | h_1)/P(e_1 | h_0)]$  (Good, 1984 [25]), and

 $F(e_1,h_1)=[\ P(e_1\mid h_1)-P(e_1\mid h_0)]/[\ P(e_1\mid h_1)+P(e_1\mid h_0)]\ (Kemeny\ and\ Oppenheim,\ 1952\ [7]).$ 

 $b^*(e_1,h_1)=[P(e_1|h_1)-P(e_1|h_0)]/\max(P(e_1|h_1), P(e_1|h_0)), (Lu, 2000 [8]).$ 

这些测度都和似然比(Likelihood Ratio,缩写为 LR)正相关,比如  $L=\log LR^+$ ,  $F=(LR^+-1)/(LR^++1)$ 【7】,因而这些测度和医学检验及疾病控制所用的评价测度兼容。虽然作者长期研究语义信息论【26-28】,但是在确证问题上站在归纳派一边。理由是:信息评价用于分类之前,而确证评价用在分类之后【8】。

虽然双方指导思想不同,但是都同意用一个样本中四个样例 $(e_1,h_1)$ ,  $(e_1,h_0)$ ,  $(e_0,h_0)$ , and  $(e_0,h_0)$ 的个数或联合概率构造确证测度【10,8】。增量派的问题是:没有区分全称判断或大前提 e->h 的证据(一个样本)和结论 h 的证据 e. 当他们说要评价 e 对 h 的支持度时,e 被理解为一个样本;而当它们使用四个样例个数的时候,e 被看做是大前提的前件。当 e 被理解为前件时,说 e0 是 e1 的否定才是有意义的。当 e 被理解为样本时,其否定 e0 是无意义的, 把一个样本 e1 放进样例 $(e_1,h_1)$ 是更加无意义的!

我们比较确证测度  $D(\mathbb{P} \mathbb{S})$  中的 confirm i)和 S 看两派确证测度的数值差异。因为

$$D(e_1,h_1)=P(h_1|e_1)-P(h_1)=P(h_1|e_1)-[P(e_1)P(h_1|e_1)+P(e_0)P(h_1|e_0)]$$

$$=[1-P(e_1)]P(h_1|e_1)-P(e_0)P(h_1|e_0)=P(e_0)S(e_1,h_1).$$
(6)

我们可以发现 D 随 P(e)改变, 而 S 和 P(e)无关, P(e)表示信源, P(h|e)信道。D 和

两者有关, 但是 S 只和 P(h)有关。而 F 和 b\*测度既独立于信源, 也独立于信宿, 只反映信道特性(信道是 P(e|h))。这就是为什么作者称 b\*为信道确证测度。

### 2.2. 用 P-T 概率框架和语义信息方法推导出信道确证测度 $b^*(e-h)$

P-T 概率框架【28】中 P 意味着概率, T 意味着真值或逻辑概率。该框架中, 谓词的真值函数就是模糊集合的隶属函数【29】,所以真值函数也在0和1之间 变化。作者只所以提出这一概率框架,是要在统计和逻辑之间搭起桥梁。

令 X 是一随机变量,表示一个实例,取值  $x \in A=\{x_0,x_1,...\}$ ; Y 是以随机变量, 表示一个标签或假设,取值  $y \in B=\{y_0,y_1,...\}$ . 香农信道可用一个条件概率矩阵  $P(y_i|x_i)$  (i=1,2,...,m; j=1,2,...,n)或一组转移概率函数  $P(y_i|x)$  (j=1,2,...,n)表示;而语 义信道可以用一个真值矩阵  $T(y_i|x_i)(i=1,2,...,j=1,2,...)$ 或一组真值函数  $T(y_i|x)$ (j=0,1,...)表示。设 A 中使  $y_i$ 为真的元素构成一个模糊子集  $\theta_i$ ,则 x 在  $\theta_i$  中的隶 属度(记为  $T(\theta_i|x)$ )就是真值函数  $T(y_i|x)$ , 即  $T(\theta_i|x)=T(y_i|x)$ .  $y_i$ 的逻辑概率就是 Zadeh 定义的模糊事件的概率【30】:

$$T(y_j) = T(\theta_j) = \sum_i P(x_i) T(\theta_j \mid x_i). \tag{7}$$

 $y_i$ 为真时, x 发生的概率是:

$$P(x \mid \theta_i) = P(x)T(\theta_i \mid x) / T(\theta_i), \tag{8}$$

 $\theta_i$ 也可以理解为模型参数,这样  $P(x \mid \theta_i)$ 就是相应  $y_i$ 的似然函数【27】。 逻辑概率和统计概率的的区别是:

- 统计概率是归一化的,而逻辑概率不是归一化的,即一般情况下  $T(\theta_0) + T(\theta_1)$
- 每个  $T(\theta_i|x)$ 的最大值是 1(对于不同的 x),而  $P(y_0|x)+P(y_1|x)+...=1$  (对于给 定 x);

我们可以用样本分布优化模型参数。比如,我们用x表示年龄,用一个logistic函数表示老年人的真值函数 T("老年人"|x) = 1/[1+exp(-b(x-a))], 我们可以用样本 优化 a 和 b.

$$y_j$$
为真提供关于  $x_i$  的语义信息是
$$I(x_i; \theta_j) = \log \frac{P(x_i | \theta_j)}{P(x_i)} = \log \frac{T(\theta_j | x_i)}{T(\theta_j)},$$
(9)

对不同x,yi提供的平均语义信息是

$$I(X; \theta_{j}) = \sum_{i} P(x_{i} \mid y_{j}) \log \frac{T(\theta_{j} \mid x_{i})}{T(\theta_{j})} = \sum_{i} P(x_{i} \mid y_{j}) \log \frac{P(x_{i} \mid \theta_{j})}{P(x_{i})}$$

$$= -\sum_{i} P(x_{i} \mid y_{j}) \log P(x_{i}) - H(X \mid \theta_{j}).$$

交叉熵:

其中  $H(X|\theta_i)$ 是交叉熵:

$$H(X \mid \theta_j) = -\sum_i P(x_i \mid y_j) \log P(x_i \mid \theta_j). \tag{11}$$

交叉熵有个重要性质: 当我们改变  $P(x \mid \theta_i)$ 使得  $P(x \mid \theta_i) = P(x \mid y_i)$ 时, $H(X \mid \theta_i)$ 达到

从等式(10)容易看出, $H(X|\theta_i)$ 达到其最小值时  $I(X;\theta_i)$ 达到其最大值。作者还证 明【27】,如果  $P(x|\theta_i)=P(x|y_i)$ ,则  $T(\theta_i|x) \propto P(y_i|x)$ 。如果对于所有 i,  $T(\theta_i|x) \propto$ 

 $P(y_i|x)$ , 我们称语义信道匹配香农信道。

为确证医学检验中的推理,我们定义  $h \in \{h_0, h_1\} = \{\bar{e}_{N_0}, h_2\} = \{\bar{e}_{N_0}, h_1\} = \{\bar{e}_{N_0}, h_2\} = \{\bar{e}_{N_0}, h_1\} = \{\bar{e}_{N_0}, h_2\} = \{\bar{e}_{N_$ 

设一个模糊谓词  $e_1(h)$ 的真值函数等于一个清晰句(其真值仅取 0 和 1 二值)和一个永真句(其真值总是 1)的线性组合, 永真句的比例是  $b_1$ '>0,清晰句的比例是 1- $b_1$ ', 则

$$T(e_1 \mid h_0) = b_1'; \quad T(e_1 \mid h_1) = b_1' + b_1 = b_1' + (1 - b_1') = 1.$$
 (12)

 $b_1$ '就是  $e_1$ -> $h_1$  的不信度. 样本优化得到的不信度  $b_1$ \*就是否证度。确证度  $b_1$ \*和否证度  $b_1$ \*的关系是  $b_1$ \*+| $b_1$ \*|=1. 然后我们可以推导出(详见【8】中 3.2 节):

$$b_1^* = b^*(e_1 \to h_1) = \frac{P(e_1 \mid h_1) - P(e_1 \mid h_0)}{\max[P(e_1 \mid h_1), P(e_1 \mid h_0)]} = \frac{LR^+ - 1}{\max(LR^+, 1)}.$$
 (13)

其中  $LR+=P(e_1|h_1)/P(h_1|e_0)$  是阳性  $e_1$  的似然比。.

假设似然函数  $P(h|e_1)$ 可以分解成等概率部分和仅含 0 和 1 的部分,我们可以推导出预测确证度

$$c_1^* = c^*(e_1 \to h_1) = \frac{P(h_1 \mid e_1) - P(h_0 \mid e_1)}{\max(P(h_1 \mid e_1), P(h_0 \mid e_1))} = \frac{2P(h_1 \mid e_1) - 1}{\max(P(h_1 \mid e_1), 1 - P(h_1 \mid e_1))}. (14)$$

测度  $b^*$ 兼容似然比(likelihood ratio),适合评价医学检验——作为手段;而  $c^*$ 适合评价结论的必然性,可以用来澄清乌鸦悖论。两者都具有对称性和归一化性质。

## 2.3. 因果推断 -- 从辛普森悖论(SP)谈起

对于例 2(关于肾结石),分组结论是对的,合并结论是错的。而对于例 3(关于吃降压药),合并结论是对的。这是因为  $P(y_1|x_1)$ 和  $P(y_1|x_0)$ 并不能很好反映因果关系。除了观察数据,即样本的联合概率分布 P(y,x,g),我们还需要假设数据背后的因果结构或因果图【3】。

假设除了原因 x 和结果 y, 还存在第三未知变量 u. 图 2(a),(b), (c)分别表示例 2, 例 3, 例 4 中的因果关系。图 2(a)显示了例 2 的因果结构,其中 u(表示肾结石大小)是混杂因素(confounder),它影响 x 也影响 y。图 2(b)描述了例 3 的因果结构,其中 u(表示血压)是中间因素(mediator),它受 x 影响,也影响 y。在图 2(c)中,u 可以解释为混杂因素,也可以理解为中间因素;如果视角不同,因果关系就不同, $P(y_1|do(x))$ 也就不同。在所有例子中,我们都要用  $P(y_1|do(x))$ 取代  $P(y_1|x)$ ——如果两者不同,得到 RD,RR 和  $P_d$ 。

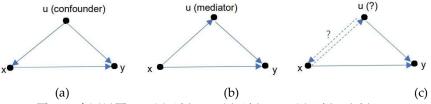


图 2 三个因果图. (a) 用于例 2; (b)用于例 3; (c) 用于例 4 和例 1.

对于 u 是中间因素的例子,我们接收合并结论就行了。而对于 u 是混杂因素 8 / 21

的例子,如何得到合理的 P(y|do(x))? 根据 Rubin 的潜在结果模型,我们用图 3 解释 P(y|do(x))和 P(y|x)的区别。

要发现一种原因  $x_2$  和另一种原因  $x_1$  引起的结果差异,我们需要比较两者在相同背景下导致的结果。但是现实中往往不存在其他条件不变而只有原因改变的情况。为此我们需要在想象中用  $x_2$ 取代  $x_1$ ,或用  $x_1$ 取代  $x_0$ ,看结果  $y_1$  发生的概率的变化。 如果 u 不受 x 影响, $g_1$  和  $g_2$  的成员数量就应该不随 x 变化,如图 3 所示。解决办法是,用 P(g)取代 P(g|x)做加权运算。这样就能保证合并结论和分组结论一致,不存在悖论。

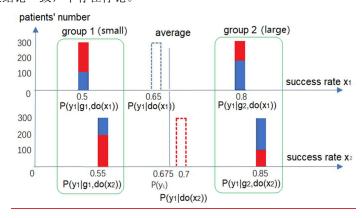


图 3. 通过更改加权系数消除辛普森悖论(混杂因素存在时)。用  $P(g_k)$ 代替  $P(g_k|x_i)$ 后 (k=1,2; i=1,2) 做加权运算,整体结论和分组结论一致。  $x_2$  的平均成功率  $P(y_1|do(x_2))=0.7$  高于  $x_1$  的平均成功率  $P(y_1|do(x_2))=0.65$ .

虽然  $P(x_0)+P(x_1)=1$  成立,但是  $P(do(x_1)))+P(do(x_0))=1$  是无意义的。这也是为什么 Rubin 强调  $P(y^{x_1})$ 和  $P(y^{x_0})$ ——即 P(y|do(x))和  $P(y|do(x_0))$ ——在本质上仍然是边际概率,而不是条件概率。

Rubin 用 P(g)取代 P(g|x)的依据是【2】:对于每一组,比如  $g_1$ ,被  $x_1$ 和  $x_2$ 干 预的两组成员(比如病人)是可交换的,分在  $x_1$ 组,得到结果的条件概率就是  $P(y_1|g_1,x_1)$ ,分在  $x_2$ 组,得到结条件概率就是  $P(y_1|g_1,x_2)$ .  $P(g_1|x_1)$ 和  $P(g_1|x_2)$ 不同是由于数据缺失一半,我们在想象中补齐缺失数据就行了,见图 3。

如果 u 是中间因素,如图 2(b)所示,在  $x_1$  的作用下, $g_1$  的成员可能进入  $g_2$ ,或相反。这时候  $P(g|x_0)$ 和  $P(g|x_1)$ 自然不同,不需要用 P(g)取代它们。我们可以直接令  $P(y_1|do(x_1))=P(y_1|x_1)$ 且  $P(y_1|do(x_0))=P(y_1|x_0)$ , 并接受合并结论。

#### 2.4 因果概率测度

在 Rubin 和 Greenland 的 1989 的文章【13】中,

$$P(t) = [R(t) - 1] / R(t)$$
(15)

被解释为因果关系概率,其中 t 是病人曝光到有害环境的年龄; R(t)是年龄相关的感染率,即感染的人数除以未感染人数。 如果  $y_1$ 表示感染, $x_1$ 表示曝光(到有害环境), $x_0$ 表示没曝光,则有  $R(t) = P(y_1 \mid do(x_1), t) / P(y_1 \mid do(x_0), t)$ . 其下限是 0,

因为概率不能是负的。如果忽略 t 的变化,考虑到下限,我们可以把因果关系概率写成

$$P_d = \max(0, \frac{R-1}{R}) = \max(0, \frac{P(y_1 \mid do(x_1) - P(y_1 \mid do(x_0)))}{P(y_1 \mid do(x_1))}).$$
(16)

Pearl 使用"PN"代表  $P_d$ ,并把 PN 解释为 the probability of necessity 【3】(原 因必要性的概率).

 $P_d$ 和确证度  $b^*$ 【8】很相似. 主要差别是:  $b^*$ 在-1 和 1 之间变化.

Robert van Rooij 和 Katrin Schulz 【31】 争论说形如"如果 x, 那么 y"的条件句是可主张的,只有下面测度是高的:

$$\Delta * P_x^y = \frac{P(y_1 \mid x_1) - P(y_1 \mid x_0)}{1 - P(y_1 \mid x_0)}.$$
(17)

这一测度和测度 Z 相似.  $P_a$  和  $\Delta^*P_{x''}$  差别:  $P_a$  像  $b^*$ ,对反例的比例  $P(y_1|x_0)$  敏感,而  $\Delta^*P_{x''}$  不敏感。表 1 显示了他们的差别。

Table 1 比较  $P_d$  和  $\Delta^* P_{xy}$ 

	$P(y_1 x_1)$	$P(y_1 x_0)$	$P_d$	$\Delta^* P_{x^y}$
两者差别小	0.9	0.8	0.11	0.5
无反例	0.2	0	1	0.2

David E Over 等人【32】支持 Ramsey 检验假设——一个自然语言条件句的 主观概率 P(if p then q)就是条件主观概率  $P(q \mid p)$ . 这一测度和 confirm f 【5】相 同。.

作者建议【8】我们应当区分关于 x->y 的两类确证测度。一是反映  $x(相对于x_0)$ 的必要性,二是反映 y 的不可避免性。P(y|x)作为后者是好的,但是作为前者并不好。前者应该独立于 P(x)和 P(y).  $P_d$  就是这样的测度.

然而, $P_a$ 有个重要问题,在 x 不影响 y 时  $P_a$ =0 是,这是合理的。 但是,如果 x 对 y 的影响是负的,如疫苗影响感染, $P_a$  最好小于 0. 因此,要度量因果强度,我们需要确证测度(在-1 和 1 之间变化)而不是概率测度(在 0 和 1 之间变化)。

### 3. 方法

#### 3.1 定义因果后验概率

为避免把关联关系看做因果关系,我们首先说明什么样的后验概率反映因果关系。后验概率和条件概率通常不加区分。但是 Rubin 强调  $P(y^*)$ 并不是条件概率,它仍然是边际概率。为区分  $P(y^*)$ 和 P(y),我们称  $P(y^*)$ ,即  $P(y^*|$  do(x)),是因果后验概率(Causal Posterior Probability,缩写为 CPP). 什么样的后验概率是CPP? 我们使用下面例子讨论。

例 1. 我们以人口年龄分布为例说明。设 z 是年龄,人口年龄密度分布是 P(z), $z \ge 60$  被称为老年人,即, $P(y_1|z)=1$  当  $z \ge z_0=60$ .

老年人的标签是  $y_1$ ="老年人",非老年人的标签是  $y_0$ . 老年人的概率是 10 / 21

$$P(y_1) = \sum_{\text{all } z} p(z)P(y_1 \mid z) = \sum_{z > 60} p(z)$$
(18)

设  $x_1$  表示改进的医疗条件,则 P(z)在未来会变为  $P(z^{x_1})$ . 则提供  $x_1$  的一短时间后 (比如 10 年),老年人的概率变为

$$P(y_1^{x_1}) = P(y_1 \mid do(x_1)) = \sum_{z > so} p(z \mid do(x_1)),$$
(19)

设 $x_0$ 为已有医疗条件,则有

$$P(y_1 \mid do(x_0)) = \sum_{z > 60} p(z \mid do(x_0)).$$
 (20)

和上面类似的例子还有:

- 一种药物(x)是否能降低血压, 血糖, 血脂或尿酸(z), z 降到一定幅度  $z_0$  算是治疗成功( $y_1$ );
- 一种手术( $x_1$ )是否延长病人存活时间(z), z 长到一定时间  $z_0$  算是手术成功( $y_1$ );
- 一种保健品( $x_1$ )是否能延长人的寿命(z), 吃了保健药的人的z超过平均寿命  $z_0$ 的 3%就算有效  $y_1$ ;
- 一种肥料(x₁)是否增加谷物产量(z), z 增加到一定幅度 z₀算是丰收(y₁);
- 一种工艺  $x_1$  是否能缩小产品尺寸的误差 z, 误差小到一定范围( $z_0$ )就算合格( $y_1$ );

从上面的例子看,一种条件是否可以作为因果关系的原因,就在于它是否能引起客观结果 z 的概率分布 p(z)变化,而不在于引起结果 y 的概率分布 P(y|.)的变化,因为 P(y|.)也可以随划分边界  $z_0$ 变化。比如老年人的划分边界从  $z_0$ =60 变为  $z_0$ '=65,则  $y_1$  的概率  $P(y_1|.)$ 会变为  $P(y_1|do)$ (改变  $z_0$  为  $z_0$ '))。好像这种变化似乎也反映因果关系。但是笔者以为,这种变化是由于数学的原因,不反映我们要研究的因果关系。为此,我们需要更具体定义 CPP。

定义 1: 设随机变量 Z 取值是  $z \in \{z_1, z_2, ...\}$ , P(z)表示客观结果的概率分布;随机变量 Y 的取值是  $y \in \{y_0, y_1\}$ , y 表示结果,即 z 的分类标签. 原因是  $x \in \{x_0, x_1\}$  或 $\{x_1, x_2\}$ 。如果用  $x_1$ 取代  $x_0$ (或用  $x_2$ 取代  $x_1$ )能引起纯客观结果 z 的概率分布 P(z) 变化,我们称 x 是原因,称  $P(z^x)$ 是 CPP 分布,称  $P(y^x) = P(y^y \mid do(x))$ 是 CPP。

按照上面定义,给定 $y_1$ 时,z的条件概率分布 $P(z|y_1)$ 不是因果条件概率,因为z的概率分布不随 $y_1$ 变化。

假设  $x_1$  是疫苗, $y_1$  是感染, $e_1$  是检验阳性。 $P(y_1|x_1)$ 或  $P(y_1|do(x_1)$ 是 CPP,而  $P(y_1|e_1)$ 不是 CPP。 $y_1$  可以理解为通过最好检验得到的结论, $e_1$  是一种普通检验, $P(y_1|e_1)$ 是对  $y_1$  的预测。因为 e 的提供并不改变 P(z)及最好检验结论,所以  $P(y_1|e_1)$ 不是 CPP。

## 3.2 使用 $x_2/x_1=>y_1$ 比较两种原因对结果的影响

在关联关系中,  $x_0$  是  $x_1$  的非, 两者是互补的。但是在因果关系中,  $x_1$  是  $x_0$  的替代者。考虑吃药治病,  $x_0$  是不吃药, 吃不同药可用  $x_1$ ,  $x_2$ ...表示。后者每一个都是  $x_0$  的可能替代者而不是非。特别值得注意的是, 在某些情况下,  $x_1$  可能包括  $x_0$ (见 4.3 节).

当我们比较 $x_2$ 相对于 $x_1$ 引起的效果时,用" $x_2=>y_1$ "表示因果关系是不明了的。

因此,作者建议我们最好用" $x_2/x_1=>y_1$ "(表示"用  $x_2$ 代替  $x_1$ 会增加  $y_1$ ) 取代" $x_2=>y_1$ "。 这样做有两个理由:

- 便于表达确证度的对称性(见 3.3 节), 即  $Cc(x_2/x_1=>y_1)=-Cc(x_1/x_2=>y_1)$ ;
- 强调 x<sub>1</sub>和 x<sub>2</sub>不是互补的,而是相互取代的,以便更容易澄清 SP.

对于存在混杂的情况,比如 Example 2,如果在想象中用  $x_2$  取代  $x_1$ ,我们就能容易理解为什么每一组的病人数目是不变的, 即  $P(g|x_1)=P(g|x_2)=P(g)$ 。因为每个人的肿瘤大小不会因为使用另外一种手术而改变。,

而在例 3 中,u 是中间因素,每组(高或低血压)人数也受吃降压药( $x_1$ )影响,当我们用  $x_1$  取代  $x_0$ (不吃药)时, $P(g|x_1)\neq P(g|x_0)\neq P(g)$ 是有理由的,所以加权系数无需调整,我们直接令  $P(y_1|do(x))=P(y_1|x)$ 就行了。

# 3.3 用语义信息方法推导因果确证测度 Cc

我们以  $x_1 => y_1$  为例推导因果确证测度 Cc. 如果我们需要比较两个原因 xi 和 xk 引起的结果,我们可以假设其中一个是缺省的如 xo.

令  $s_1="x_1=>y_1"$ ,  $s_0="x_0=>y_0"$ . 我们假设  $s_1$  包含的可信部分比例是  $b_1$ , 不可信部分比例是  $b_1'$ . 两个比例之间的关系是  $b_1'+|b_1|=1$ . 首先我们假设  $b_1>0$ , 于是  $b_1=1-b_1'$ .  $S_1$  的两个真值如表 2 所示。

表 2 因果陈述句 so="xo=>yo"和 s1="x1=>y1"的真值

$T(s \mid x)$	$\chi_0$	$\chi_1$	
$s_0 = "x_0 = y_0"$	1	$b_0$ '	
$s_1 = "x_1/x_0 = > y_1"$	$b_1'$	1	

 $T(s_1|x_1/x_0)=1$  意味着样例 $(x_1,y_1)$ 使  $s_1$  为真;  $T(s_1|x_0)=b_1'$  是反例 $(x_0,y_1)$ 的真值,也是  $s_1$ 的不信度。

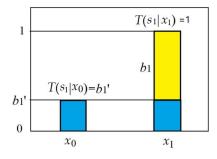


图 4. 真值函数  $T(s_1|x)$ 包括可信部分(比例是  $b_1$ )和不可信部分(比例是  $b_1$ ).

通过样本优化的可信度——用最大语义信息或最小交叉熵准则——就是确证度,记为  $Cc_1 = Cc(x_1/x_0=>y_1) = b_1^*$ .

s1的逻辑概率是

$$T(s_1) = P(x_1) + P(x_0) b_1'. (21).$$

根据 s1和 y1预测的 x1的概率是:

12 / 21

$$P(x_1 \mid \theta_1) = \frac{T(s_1 \mid x_1)P(x_1)}{T(s_1)} = \frac{P(x_1)}{P(x_1) + P(x_0)b_1},$$
(22)

其中  $\theta_1$  是表示真值函数  $T(s_1|x)$ 的参数。

y1和 s1传递的关于 x 的平均语义信息是

$$I(X; \theta_1) = \sum_{i} P(x_i \mid y_1) \log \frac{P(x_i \mid \theta_1)}{P(x_i)} = -\sum_{i} P(x_i \mid y_1) \log P(x_i) - H(X \mid \theta_1), \quad (23)$$

其中  $H(X|\theta_1)$ 是交叉熵。这里假设样本分布 P(x,y)是经过调整的,即 P(y|x)=P(y|do(x)). 根据交叉熵的性质, $H(X|\theta_i)$  达到其最小值且  $I(X;\theta_i)$ 达到其最大值,当  $P(x|\theta_1)=P(x|y_1)$ 时,即

$$P(x_0 \mid \theta_1) = \frac{P(x_0)b_1'}{P(x_1) + b_1'P(x_0)} = P(x_0 \mid y_1),$$

$$P(x_1 \mid \theta_1) = \frac{P(x_1)}{P(x_1) + b_1'P(x_0)} = P(x_1 \mid y_1).$$
(24)

从上面两个等式, 可以得到

$$\frac{P(x_0 \mid y_1)}{P(x_1 \mid y_1)} = \frac{P(x_0)}{P(x_1)} b_1', \tag{25}$$

令

$$m(x_i, y_j) = \frac{P(y_j \mid x_i)}{P(y_i)} = \frac{P(x_i, y_j)}{P(x_i)P(y_i)}, \quad i = 0, 1; j = 0, 1. \quad (26)$$

它表示  $x_i$ 和  $y_j$ 的相关度,独立于 P(x)和 P(y)——不像  $P(xi, y_j)$ . 从上面两个等式,我们得到  $s_i$  被样本优化的不信度,即否证度:

$$b_1'^* = m(x_0, y_1)/m(x_1, y_1). \tag{27}$$

然后得到 s1 的确证度:

$$b_1^* = \frac{m(x_1, y_1) - m(x_0, y_1)}{m(x_1, y_1)}.$$
 (28)

在上面公式中,我们假定  $b_1$ \*>1, 因此  $m(x_1, y_1) \ge m(x_0, y_1)$ 。 如果  $m(x_1, y_1) < m(x_0, y_1)$ , $b_1$ \*应该是负的。于是有:

$$b_1^* = -(1 - b_1^{*}) = -(1 - \frac{m(x_1, y_1)}{m(x_0, y_1)}) = \frac{m(x_1, y_1) - m(x_0, y_1)}{m(x_0, y_1)}.$$
 (29)

结合上面两个公式,我们得到确证测度:

$$Cc(x_1/x_0 \Rightarrow y_1) = b_1^* = \frac{m(x_1, y_1) - m(x_0, y_1)}{\max(m(x_1, y_1), m(x_0, |y_1))}.$$
 (30)

因为 $P(y_i|x_i) = m(x_i, y_i)P(y_i)$ , 我们也有

$$b_1' *= P(x_0 | y_1)/P(x_1 | y_1),$$
 (31)

$$Cc(x_1/x_0 \Rightarrow y_1) = b_1^* = \frac{P(y_1|x_1) - P(y_1|x_0)}{\max(P(y_1|x_1), P(y_1|x_0))} = \frac{R-1}{\max(R, 1)},$$
 (32)

其中  $R = P(y_1|x_1) / P(y_1|x_0)$  是相对风险或似然比——它也用来表达  $P_d$ .

测度 Cc 具有归一化性质和对称性,因为  $m(x_0,y_1)=0$  是 Cc=1,  $m(x_1,y_1)=0$  时 Cc=-1。并且 Cc 有原因对称性,因为:

13 / 21

$$Cc(x_{0} / x_{1} \Rightarrow y_{1}) = \frac{m(x_{0}, y_{1}) - m(x_{1}, y_{1})}{\max(m(x_{0}, y_{1}), m(x_{1}, y_{1}))}$$

$$= -\frac{m(x_{1}, y_{1}) - m(x_{0}, y_{1})}{\max(m(x_{1}, y_{1}), m(x_{0}, |y_{1}))} = -Cc(x_{1} / x_{0} \Rightarrow y_{1}).$$
(33)

用类似方式,令概率分布 P(y|x1) 是等概率分布和 0-1 概率分布的组合,我么得到另一个因果确证测度:

$$C_{e}(x_{1}/x_{0} \Rightarrow y_{1}) = \frac{P(y_{1}|x_{1}) - P(y_{0}|x_{1})}{\max(P(y_{1}|x_{1}), P(y_{0}|x_{1}))} = \frac{2P(y_{1}|x_{1}) - 1}{\max(P(y_{1}|x_{1}), 1 - P(y_{1}|x_{1}))}.$$
 (34)

该测度可看做贝叶斯确证测度  $c^*(e_1-h_1)$  【8】的推广. Ce 随  $f(h_1, e_1)=P(h_1 \mid e_1)$  (Fitelson 等人 【5,32】使用的)单调增加,但是它有归一化性质和对称性(结果对称性),即:

$$Ce(x_1=>y_1) = -Ce(x_1=>y_0).$$
 (35)

### 3.4. 因果确证测度 Cc 和 Ce 用于概率预测

从  $y_1, b_1*$ 和 P(x), 我们能得到关于 x 的概率预测:

$$P(x_1 \mid \theta_1) = \frac{P(x_i)}{P(x_1) + b_1^{*} P(x_0)}, \ P(x_0 \mid \theta_1) = \frac{P(x_0) b_1^{*}}{P(x_1) + b_1^{*} P(x_0)}, \quad (36)$$

其中  $b_1$ \*>0,  $\theta_i$  表示  $y_i$ 和  $b_i$ \*.

仿照使用贝叶斯预测确证度  $c^*$ 做概率预测【8】,我们也能利用 Ce 做概率预测。比如令  $Ce_1=Ce(x_1=>y_1)$ ,给定  $x_1$  和  $Ce_1>0$ ,我们得到:

$$P(y_1 | \theta_{x_1}) = 1/(2 - Ce_1),$$
 (37)

其中  $\theta_{x1}$ 代表  $x_1$ 和  $Ce_1$ .

给定语义信道——由两个可信度确定(参看表 2), 根据等式(32), 我们能得到相应的香农信道:

$$P(y_{1} | x_{1}) = \frac{1 - b_{0}'}{1 - b_{1}' b_{0}'}, \ P(y_{0} | x_{0}) = \frac{1 - b_{1}'}{1 - b_{1}' b_{0}'},$$

$$P(y_{0} | x_{1}) = 1 - P(y_{1} | x_{1}), \ P(y_{1} | x_{0}) = 1 - P(y_{0} | x_{0}).$$
(38)

### 4. 结果

### 4.1. 关于肾结石治疗的例子

表 3 显示了带有详细数据的例 2(关于肾结石治疗)【15】。初始数据来自【16】。 其中 \*% 意思是成功率,它后面的数字是病人数。石头大小是混杂因素。来自每一组(按结石大小分)的结论是治疗  $x_2$  好过治疗  $x_1$  【15】中治疗 A 是这里的  $x_2$ ,治疗 B 是这里的  $x_1$ ; 而根据平均成功率( $P(y_1|x_1)$ =0.83 和  $P(y_1|x_2)$ =0.78)的结论却是  $x_1$  好过  $x_2$ .治疗  $x_1$ 和  $x_2$ 那个更好?这里存在悖论。

我们用 P(g)而不是  $P(g|x_1)$ 和  $P(g|x_2)$ 做加权系数得到  $P(y_1|do(x_1))$ 和  $P(y_1|do(x_2/x_1))$ . 用  $P(y_1|do(x_1))$ 和  $P(y_1|do(x_2/x_1))$ 代替  $P(y_1|x_1)$ 和  $P(y_1|x_2)$ 后,整体结论也是  $x_2$  好过  $x_1$ 。

### 表 3. 比较两种治疗的成功率(y1意思是成功)

14 / 21

	Treat. x1	Treat. x2	Number	P(g) or $Cc$
Small stones (g1)	87%/270	93%/87*	357	0.51
Large stones (g2)	69%/80	73%/263	343	0.49
Overall	83%/350	78%/350	700	
$P(y_1   x)$	0.83	0.78		$[(P(y_1 x_2) - P(y_1 x_1)] / P(y_1 x_2) = -0.064$
$P(y_1   do(x))$	0.78	0.83		$Cc_1 = Cc(x_2/x_1 => y_1) = 0.06$
$P(y_0 \mid do(x))$	0.22	0.17		$Cc_0 = Cc(x_1/x_2 => y_0) = 0.23$

<sup>\*&</sup>quot;87%/270" 意思是: 成功率是87%, 小组人数是270.

上面我们用  $x_1$  作为缺省治疗,因果确证度  $Cc(x_2/x_1=>y_1)$  是 0.06. 相反,如果我们用  $x_2$  作为缺省的治疗,则得到  $Cc(x_1/x_2=>y_1)=-0.06$ 。 使用 Cc,我们不用担心  $P(y_1|do(x_1))$ 和  $P(y_1|do(x_2))$ 那个更大;而使用  $P_a$ ,我们必须在计算  $P_a$ 之前就考虑好。

我们用增量派的确证测度  $D(e_1, h_1)$ 比较  $x_1$  和  $x_2$ ,获得

- $P(y_1) = P(x_1)P(y_1|x_1) + P(x_2)P(y_1|x_2) = 0.805$ ,
- $P(y_1 | x_2, g_1) P(y_1) = 0.93 0.805 > 0$ ,
- $P(y_1 | x_2, g_2) P(y_1) = 0.73 0.805 < 0$ , and
- $D(x_1, y_1) = P(y_1 | x_2) P(y_1) = 0.83 0.805 > 0.$
- $D(x_1, y_2) = P(y_1 | x_2) P(y_1) = 0.78 0.805 < 0.$

结果意味着  $x_1$  好过  $x_2$ . 看起来没有悖论,这是因为当我们使用  $D(x_1,y_1)$ 的时候,我们已经回避了悖论而不是消除了悖论。

我们用这个例子检验等式(38),从两个否证度得到的香农信道 P(y|x)和表 3 下面两行中的 P(y|do(x))相同.

## 4.2. 一个例子: 消除关于 COVID-19 的辛普森悖论

表 4 显示带有详细数据的例 4, 其中关于 COVID-19 的病死率数据【33】来自美国疾控中心 2022 年 7 月 2 日的网页 (数据只包括报告的病例,实际的病例应更多,病死率应更低). 表中  $x_1$  意思是美国 NH 白人, $x_2$ 是其他人。 $P(y_1|x_1,g)$ 和  $P(y_1|x_2,g)$  是每一年龄组中  $x_1$  和  $x_2$  的病死率. 从附录 I 可以找到原始数据和中间结果.

表 4. 关于 COVID-19 非西班牙裔美国人(x1)和其他人(x2)的各年龄组病死率数据

Age Group (g)	$P(x_1 g)$	P(g)	$P(y_1 \mid x_1, g)$	$P(g \mid x_1)$	$P(y_1   x_2, g)$	$P(g \mid x_2)$
0-4 Years	44.200	0.041	0.0002	0.0349	0.0002	0.0480
5-11 Years	44.200	0.078	0.0001	0.0659	0.0001	0.0907
12-15 Years	46.300	0.052	0.0001	0.0458	0.0001	0.0578
16-17 Years	48.700	0.029	0.0001	0.0268	0.0002	0.0307
18-29 Years	48.700	0.223	0.0004	0.2081	0.0006	0.2388
30-39 Years	49.300	0.178	0.0011	0.1681	0.0019	0.1883
40-49 Years	51.000	0.146	0.0030	0.1427	0.0048	0.1493
50-64 Years	59.100	0.163	0.0102	0.1843	0.0144	0.1389
65-74 Years	67.300	0.055	0.0333	0.0704	0.0457	0.0373
75-84 Years	72.900	0.025	0.0762	0.0356	0.0938	0.0144
85+ Years	76.300	0.012	0.1606	0.0173	0.1751	0.0059
sum		1	1	1	1	1

表 5 显示了加权系数改变前后的整体(平均)病死率和风险度。

表 5. 比较 NH 美国人和其他美国人的病死率和风险度——加权系数调整前后

	NH 白人(x1) 病死率	其他美国人(x2)病死率	风险度
$P(y_1   x)$	1.04	0.73	$P_d = Cc = (R - 1) / R = 0.30$
$P(y_1   do(x))$	0.80	1.05	$P_d = 0$ ; $Cc(x_1/x_2=>y_1) = -0.28$

 $<sup>*</sup>R = P(y_1 | x_1)/P(y_1 | x_2).$ 

从表 4 我们能发现不同年龄组的 NH 白人病死率低于其他人种,而对于所有年龄组的整体(平均)平均病死率(见表 5),则是 NH 白人的病死率( $P(y_1|x_1)=1.04$ ) 高于其他种族的病死率( $P(y_1|x_2)=0.73$ )。在用 P(g)取代 P(g|x)作为加权系数后,则 NH 白人整体病死率  $P(y_1|do(x_1))=0.80$  也低于其他种族的病死率  $P(y_1|do(x_2))=1.05$ .

我们仿照 Fitelson 用  $D(x_1, y_1)$ 评估风险,则平均病死率是 0.97 (见相同网站 [33]). 我们得到:

$$D(NH 白人, 死亡) = P(y_1 | x_1) - P(y_1) = 1.04 - 0.97 = 0.07,$$
  
 $D(其他人, 死亡) = P(y_1 | x_2) - P(y_1) = 0.73 - 0.97 = -0.14,$ 

这意味着 NH 白人面临更高风险。

## 4.3. COVID-19: 疫苗对病死率和死亡率的负影响

用因果概率测度  $P_a=(R-1)/R$  不便于度量"疫苗=>感染"和"疫苗=>死亡"的"概率",因为  $P_a$  作为概率,其最小值是 0. 而使用因果确证测度 Cc 没有问题,因为 Cc 可以是负的。

表 6 显示了从美国疾病控制和预防中心网站可以得到的数据【34】和两个因果确证度。

表 6. 一周内 (6月 20-26) 每 10000 万人口中感染和死亡数 (5岁以上) 及确证度

	Unvaccinated (x0)	Vaccinated (x1)	Cc
Cases	512.6	189.5	$Cc(x_1/x_0=>y_1)=-0.63$
Deaths	1.89	0.34	$Cc(x_1/x_0)=>y_1)=-0.79$
Mortality rate	0.001	0.00018	

负的确证度-0.63 和-0.79 度意味着和未打疫苗的人相比,打过疫苗的人感染数减少了 63%, 因感染而死亡的人数减少了 79%。

要知道 COVID-19 对人口死亡率的影响,我们需要比较相同时间内(比如 1年)普通死亡数和新的死亡数。美国人均寿命是79岁,年死亡率大约是1/79=0.013. 从表 6 可以得到: COVID-19 导致的年死亡率是 0.1%(对于没打疫苗的人)和 0.018%(对于打了疫苗的人).

因为 COVID-19 死亡的人,也可能因为普通原因在同一年死亡。所以新的死亡率小于两个死亡率相加。假设等于两者相加,则新的死亡率  $P(y_1|x_1)$ 是 0.013+0.001=0.014(对于没打疫苗的人)和 0.013+0.00018=0.01318(对于打了疫苗的人). 表 6显示了 COVID-19 对死亡率的影响。其中假设  $P(y_1|x)=P(y_1|do(x))$ 。

表 7. COVID-19 对人口死亡率的影响

	Mortality rate $P(y_1 x)$	Unvaccinated	Vaccinated
x <sub>0</sub> : common reasons	$P(y_1   x_0)$	0.013	0.013

x1: x0 plus COVID-19	$P(y_1 \mid x_1)$	0.014	0.01318
$Cc_1=Cc(x_1/x_0=>y_1)$		0.07	0.014

Cc=0.07 和 Cc=0.014 意味着按照目前病毒杀伤力,死亡人数中,没打疫苗的人将有 7%、打了疫苗的人将有 1.4%是由于 COVID-19.

上面结果没有考虑疫苗副作用——或许会导致慢性死亡。

## 5. 讨论

#### 5.1 理论上的理由: 为什么说 $P_a$ 和 Cc 能更好反映因果关系强度?

我们称  $m(xi, y_i)$  (i=0,1; j=0,1)是概率相关矩阵,它不是对称的。虽然从计算的角度看,先有 P(x,y)后有 m(x,y),但是从存在的角度或因果的角度看,是先有 m(x,y),后有 P(x,y)。即给定 P(x),m(x,y)只允许特定的 P(y)产生。

用 m(x, y)也可以做概率预测(如同用贝叶斯公式做概率预测):

$$P(y | x_1) = P(y)m(x_1, y) / m(x_1), \quad m(x_1) = \sum_{y} P(y)m(x_1, y),$$

$$P(x | y_1) = P(x)m(x, y_1) / m(y_1), \quad m(y_1) = \sum_{y} P(x)m(x, y_1).$$
(39)

从等式 (25)-(27)可见, $P_a$ 和 Cc 只取决于 m(x, y),独立于 P(x)和 P(y). 两个否证度确定  $b_1$ \*\*和  $b_0$ \*\*确定了一个语义信道和一个香农信道(见 Table 3 下面两行)。用两个否证度  $b_1$ \*\*和  $b_0$ \*\*可以做概率预测,见公式 (36)。所以,两个确证度  $Cc_1=b_1$ \* 和  $Cc_0=b_0$ \*表示的是从 x 到 y 的约束关系——即因果关系——的强度。测度  $P_a$ 和 Cc 类似;而 D 和  $\Delta^*P_x$ \*都和 P(x)有关,所以它们并不单纯反映因果关系。

比如:考虑疫苗对 COVID-19 病死率的影响(见表 7),  $P_a$ 和  $C_c$  和疫苗覆盖率  $P(x_1)$ 无关,而测度  $\Delta^*P_{x_2}$  明显地和 P(x)有关;测度 D 和 P(y)有关,因而也和  $P(x_1)$ 有关。从一个地方得到的  $P_a$ 和  $C_c$ 1 也适合评价同样病毒在其他地方对致死率的影响。而用  $\Delta^*P_{x_2}$ 和 D 不能通用,因为疫苗覆盖率  $P(x_1)$ 在不同地方是不同的。

按照 Bayesian 确证的增量派观点, $P(y_1)$ 是先验概率, $P(y_1|x)$ 是在它基础上的增量。但是将测度 D 用于因果确证时, $P(y_1)$ 包括干预后的数据,因而不再表示先验概率了. 这在理论上是一个致命问题。

另外,Cc 作为归纳的结果,能反映模糊大前提,用于概率预测,而 D 和  $\Delta^*P_{x^y}$  不能。

## 5.2. 实践上的理由: 为什么说 $P_a$ 和 Cc 比 D 更好

4.1 和 4.2 节两个计算实例支持这样的结论: 测度  $P_a$ 和  $C_c$  比 D(或 comfirm i) 更好。理由如下。

#### 5.2.1. 和测度 D 比, Pa和 Cc 有明确含义

 $C_{c_1}=C_c(x_1/x_0=>y_1)$ 反映的是,造成同样后果  $y_1$ , 其中有多少是因为用  $x_1$  取代了  $x_0$ 。 Pd 类似。比如:表 6 显示了,按照当时的病毒杀伤力,对于打了疫苗的人,COVID-19 使人口死亡率将从 0.013(1.3%)升高到 0.01318(1.318%),因果确证度 17 / 21

从而是  $P_a=C_{C_1}=0.014$ ,意思是死亡者中有 0.014 倍(即 1.4%)的人会是因为 COVID-19。

但是 D 和  $\Delta^* P_{xy}$  的含义就不那么清楚。

和测度 RD 不同,  $P_a$ 和 Cc 反映的是相对风险,它更能让人看清原因影响结果的程度。 很多人光看 COVID-19 一年给一个国家带来上百万人口的死亡,就觉得数目很大,风险很大,但是使用  $P_a$ 和 Cc 评估,这个死亡率和正常年份比,超出的比例并不很大(见表 6)。另一方面,过度抗疫带来的经济损失也很大。因此, $P_a$ 和 CC 有助于决策者在减少死亡和减少经济损失之间权衡利弊。

## 5.2.2. Pa和 Cc 消除了混杂因素的影响

有混杂因素存在时,如 4.1 节所示,使用  $P_4$ 或 Cc,我们能消除 SP,使整体结论和分组结论一致:手术  $x_2$ 好过  $x_1$ . 如果使用  $D(x_1,y_1)$ 评价因果关系,虽然回避了 SP,但是结论却不合常理。原因是忽略了肾结石大小不同的手术难度。按照总体成功率高的标准,如果一个医院只接受容易治疗的病人,其总体成功率就高,就是好医院。然而,这是不合常理的。

## 5.2.3. Pa和 Cc 允许我们从不同的角度看第三因素 u

如 4.3 节所示。如果我们认为人种导致长寿,则我们可以把寿命做是中介原因,然后直接接受整体结论(NH 白人比其他人种病死率高)就行了,如 Fitelson使用测度 D 得到的结论。但是,如果我们认为长寿不是因为人种,则寿命就是混杂因素,用  $P_a$ 和 Cc 就能使得整体结论和各年龄组结论一致:NH 白人的病死率比其他人种的病死率低。 $P_a$ 和 Cc 允许我们从不同角度看问题,从而得到不同结论。而测度 D 只允许我们接受整体结论。

值得注意的是,结论(NH 白人的病死率比其他人种的病死率更低)很可能是因为医疗条件影响病死率。而已有数据源不包含医疗条件信息。否则,把医疗条件当做混杂因素,双方病死率应当差不多。

# 5.3. 为什么最好用 Cc 代替 Pa?

4.3 节提供了疫苗影响死亡的确证度计算——确证度是负的。负的确证度很好理解,那就是疫苗能减少死亡。但是如果使用测度  $P_a$ (因果的概率),则  $P_a$  只能取其下限 0. 虽然我们可以用  $P_a$ (不打疫苗=>死亡)代替  $P_a$ (打了疫苗=>死亡),保证  $P_a$ >0,但是这样(把打疫苗看做是缺省原因)不符合我们的思维习惯。另外, $C_C$ 还有对称性而  $P_a$ 没有。

在比较两种原因(比如两种手术)时使用  $P_a$ , 我们总要考虑  $P(y_1|x_2)$ 和  $P(y_1|x_1)$ 那个大(见 4.1 节),而使用 Cc 就不需要考虑,因为不用担心(R-1)/R<0。

数学中的相关系数就介于 1 和-1 之间。Cc 可谓概率相关系数。不同的是,前者只有一个系数,而后者有两个系数:  $Cc_1=Cc(x_1=>y_1)$ 和  $Cc_2=Cc(x_2=>y_2)$ 。

#### 5.4. 因果关系中的必要性和充分性

 $P_d$ 和 Cc 只反映原因 x 对于结果 y 的必要性,不能反映原因对于结果的充分 18 / 21

性或结果的不可避免性。另一方面,测度 f=P(y|x)和 Ce 能表示结果的不可避免性。

医学界还用优势率(Odds Rate)同时表示 x 的必要性和 y 的不可避免性【2】:

$$OR = \frac{P(y_1 \mid x_1)}{P(y_1 \mid x_0)} \times \frac{P(y_0 \mid x_0)}{P(y_1 \mid x_0)}.$$
 (40)

它实际上是两个似然比的乘积。为了同样目的,我们可以使用

$$OR_{N} = \frac{OR - 1}{\max(OR, 1)} \tag{41}$$

作为  $x_0=>y_0$  和  $x_1=>y_1$  的总的确证测度。 $OR_N$  具有归一化性质和对称性。

## 5.5. 贝叶斯确证测度 b\*和 c\* 同因果确证测度 Cc 和 Ce 之间联系

假设  $P(y_1|x)$ 已经被调整使得  $P(y_1|x) = P(y_1|do(x))$ 。 反映原因 x 必然性的因果确证测度 Cc 在数值上等于信道确证度  $b^*$ 【8】, 即:

$$Cc(x_1 => y_1) = [P(y_1 \mid x_1) - P(y_1 \mid x_0)] / \max(P(y_1 \mid x_1), P(y_1 \mid x_0)) = b^*(y_1 \to x_1).$$
(42)

但是它们的前后件是颠倒的。这意味着,如果  $x_1$ 是  $y_1$ 的原因,则  $y_1$ 是  $x_1$ 的证据。比如被 COVID-19 感染是检验阳性的原因,则检验阳性是感染的证据。

因果确证测度 Ce 表明结果的必然性,在数值上等于预测确证测度  $c^*(x_1-y_1)$  **【8】**,即:

$$Ce(x_1 = y_1) = [P(y_1 | x_1) - P(y_0 | x_1)] / \max(P(y_1 | x_1), P(y_0 | x_1)) = c^*(x_1 \to y_1).$$
(43)

两者前后件是相同的。

然而,从上面两个等式中右边的值,我们可能得不到左边的值,因为关联关系未必是因果关系。

# 6. 结论

贝叶斯确证的增量派代表人之一 Fitelson 推荐的确证测度  $D(x_1,y_1)=P(y_1|x_1)-P(y_1)$ 用于评价因果关系时不考虑混杂因素,本文说明,这不仅和 Rubin 和 Pearl 理论为代表的因果推断理论不兼容,也和流行的风险测度  $P_d(x_1=>y_1)=\max(0, (R-1)/R)$ 不兼容。用增量派的确证理论和测度,我们只能回避 SP,而不能像因果推断理论那样消除 SP 或提供对 SP 的合理解释。

另一方面,Rubin 等人使用  $P_a$ 作为因果概率, $P_a$ 比 D 更好,但是称  $P_a$ 为因果关系概率并且用概率测度度量因果关系是不适当的。如果把  $P_a$  当做因果确证测度,则它缺少一个理想确证测度该有的归一化性质和对称性。

本文通过语义信息方法推导出因果确证测度  $Cc(x_1=y_1) = (R-1) / \max(R, 1)$ 。它和作者曾提出的归纳派确证测度  $b^*$ 类似。但是其中正例比例  $P(y_1|x_1)$ 和反例比例  $P(y_1|x_0)$ 经过去混杂影响调整(调整使得 P(y|x)=P(y|do(x))). 测度 Cc 是改进的  $P_a$ ,它具有归一化性质(在-1 和 1 之间变化)和原因对称性( $Cc(x_0/x_1=>y_1) = -Cc(x_1/x_0=>y_1)$ ). 因为 Cc 可以是负的,它因而也适合评价原因对结果的抑制关系,比如疫苗对感染的抑制关系。

本文提供了一些实例,包括两种肾结石手术比较例子和 COVID-19 的病死率,死亡率和疫苗效果的例子,计算了它们的因果确证度。计算结果支持分析结论。

另外,本文还提供了反映结果必然性的因果确证测度  $Ce(x_1=>y_1)$ ,它和作者曾提出的预测确证测度  $c^*$ 类似,也具有 Normalizing 性质和对称性。

因为测度 Cc 和流行的因果推断理论相互支持,归纳派的贝叶斯确证测度(包括  $b^*$ )因而也受到流行的因果推断理论的支持。关于因果关系的哲学(归纳派确证理论)和数学(PCIT)因此更加兼容。

但是,和所有贝叶斯确证测度一样,在样本有限的情况下所得到的确证度都不是严格可靠的,我们有必要用确证度区间取代单个确证度,保留不可避免的不确定性。这项工作有待结合已有理论进一步研究。

鸣谢:作者感谢复旦大学张志林教授和长沙学院周建勇教授,因为本文得益于和他们的交流。作者也感谢两位审稿人的评论和建议——帮助作者大大改进了本文。

附录 I: 比较 NH 美国人和其他美国人的病死率数据

原始数据来自美国疾控中心网站[33]. 包含原始数据和中间结果的 Excel 文件可从下面地址下载 http://survivor99.com/lcg/cm/CFR.zip.

### 参考文献

- Rubin, D. Causal inference using potential outcomes. J. Amer. Statist. Assoc. 2005, 100, 322–331. https://doi.org/10.1198/01621450400001880.
- Hernán, M.A.; Robins, J.M. Causal Inference: What If, Chapman & Hall/CRC: Boca Raton, FL, USA, 2020
- 3. Pearl, J. Causal inference in statistics: An overview. Stat. Surv. 2009, 3, 96–146.
- 4. Geffner, H.; Rina Dechter, R.; Halpern. J.Y. (Eds.) *Probabilistic and Causal Inference: The Works of Judea Pearl, Association for Computing Machinery*, Association for Computing Machinery, New York, United States, 2021.
- Fitelson, B. Confirmation, Causation, and Simpson's Paradox. Episteme 2017, 14, 297–309. https://doi.org/10.1017/epi.2017.25.
- Carnap, R. Logical Foundations of Probability, 2nd ed.; University of Chicago Press: Chicago, IL, USA, 1962.
- 7. Kemeny, J.; Oppenheim, P. Degrees of factual support. Philos. Sci. 1952, 19, 307–324.
- 8. Lu, C. Channels' Confirmation and Predictions' Confirmation: from the Medical Test to the Raven Paradox. *Entropy* **2020**, 22, 384. https://doi.org/10.3390/e22040384.
- Greco, S.; Slowiński, R.; Szczęch, I. Properties of rule interestingness measures and alternative approaches to normalization of measures. *Inf. Sci.* 2012, 216, 1–16.
- Crupi, V.; Tentori, K.; Gonzalez, M. On Bayesian measures of evidential support: Theoretical and empirical issues. *Philos. Sci.* 2007, 74, 229–252.
- Eells, E.; Fitelson, B. Symmetries and asymmetries in evidential support. Philos. Stud. 2002, 107, 129– 142.
- Relative Risk, Wikipedia the Free Encyclopedia. Available online: https://en.wikipedia.org/wiki/Relative\_risk (accessed on 15 August 2022).
- Robins, J.; Greenland, S. The probability of causation under a stochastic model for individual risk. Biometrics 1989, 45, 1125–1138.
- Simpson, E.H. The interpretation of interaction in contingency tables. J. R. Stat. Soc. Ser. B 1951, 13, 238–241.
- Simpson's paradox. Wikipedia the Free Encyclopedia. https://en.wikipedia.org/wiki/Simpson%27s\_paradox (accessed on 20 August 2022).
- Charig, C.R.; Webb, D.R.; Payne, S.R.; Wickham, J.E. Comparison of treatment of renal calculi by open surgery, percutaneous nephrolithotomy, and extracorporeal shockwave lithotripsy. *Br. Med. J.* (Clin. Res. Ed.) 1986, 292, 879–882. doi:10.1136/bmj.292.6524.879.
- Julious, S.A.; Mullee, M.A. Confounding and Simpson's paradox. BMJ 1994, 309, 1480–1481. doi:10.1136/bmj.309.6967.1480.

批注 [lcg1]: Font corrected

批注 [M2]: Please add the name of the publisher and their location.

批注 [lcg3R2]: I added them and deleted the link because the link is not for the full book.

批注 [M4]: Refs. [10] and [23] are same. Please check.

批注 [lcg5R4]: You are right. I deleted [23] and corrected all citing numbers.

- 18. Pedagogy, W. Simpson's Paradox. Available online: https://weapedagogy.wordpress.com/2020/01/15/5-simpsons-paradox/ (accessed on 21 August 2022).
- Mackenzie, D. Race, COVID Mortality, and Simpson's Paradox. Available online: http://causality.cs.ucla.edu/blog/index.php/category/simpsons-paradox/\_(accessed on 22 August 2022).
- Kügelgen, J.V.; Gresele, L.; Schölkopf, B. Simpson's Paradox in COVID-19 case fatality rates: A mediation analysis of age-related causal effects. *IEEE Trans. Artif. Intell.* 2021, 2, 18–27. https://doi.org/10.1109/TAI.2021.3073088.
- 21. Mortimer, H. The Logic of Induction; Prentice Hall: Paramus, NJ, USA, 1988.
- 22. Horwich, P. Probability and Evidence; Cambridge University Press: Cambridge, UK, 1982.
- 23. Christensen, D. Measuring confirmation. J. Philos. 1999, 96, 437-461.
- 24. Nozick, R. Philosophical Explanations; Clarendon: Oxford, UK, 1981.
- 25. Good, I.J. The best explicatum for weight of evidence. J. Stat. Comput. Simul. 1984, 19, 294-299.
- 26. Lu, C. A generalization of Shannon's information theory. Int. J. Gen. Syst. 1999, 28, 453-490.
- Lu, C. Semantic Information G Theory and Logical Bayesian Inference for Machine Learning. Information 2019, 10, 261. https://doi.org/10.3390/info10080261.
- Lu, C. The P–T Probability Framework for Semantic Communication, Falsification, Confirmation, and Bayesian Reasoning. *Philosophies* 2020, 5, 25. https://doi.org/10.3390/philosophies5040025.
- 29. Zadeh, L.A. Fuzzy Sets. Inf. Control 1965, 8, 338-353.
- 30. Zadeh, L.A. Probability measures of fuzzy events. J. Math. Anal. Appl. 1986, 23, 421–427.
- 31. Rooij, R.V.; Schulz, K. Conditionals, causality and conditional probability. *J. Log. Lang. Inf.* **2019**, *28*, 55–71. https://doi.org/10.1007/s10849-018-9275-5.
- 32. Over, D.E.; Hadjichristidis, C.; Jonathan St, B.T.; Evans, J.S.B.T.; Handley, D.J.; Sloman, S.A. The probability of causal conditionals. *Cogn. Psychol.* **2007**, *54*, 62-97.
- Demographic Trends of COVID-19 Cases and Deaths in the US Reported to CDC. The Website of the US CDC. Available online: https://covid.cdc.gov/covid-data-tracker/#demographics (accessed on 10 September 2022).
- Rates of COVID-19 Cases and Deaths by Vaccination Status. The Website of US CDC. Available
  online: https://covid.cdc.gov/covid-data-tracker/#rates-by-vaccine-status (accessed on 8 September
  2022)

作者更多文章见: <u>http://survivor99.com</u>

批注 [lcg6]: Thank you for your corrections! It must be much better.