

# 标题 title

作者 *author*

2024 年 7 月 24 日

# 前言

# 目录

前言	i
第一部分 AI 的逻辑	1
第一章 合情推理	2
§1.1 命题逻辑的演绎推理	3
§1.2 合情推理的数学模型	8
1.2.1 合情推理的基本假设，似然	9
1.2.2 似然与概率	12
1.2.3 先验与基率谬误	14
§1.3 合情推理的归纳强论证	15
1.3.1 归纳强论证	15
1.3.2 有效论证和归纳强论证的比较	18
§1.4 先验模型的存在性	21
§1.5 章末注记	23
第二章 Markov 链与决策	22
§2.1 Markov 链	22
§2.2 Markov 奖励过程 (MRP)	26
§2.3 Markov 决策过程 (MDP)	29
§2.4 隐 Markov 模型 (HMM)	33
2.4.1 评估问题	34
2.4.2 解释问题	35

<b>第二部分 信息与数据</b>	<b>37</b>
<b>第三章 信息论基础</b>	<b>38</b>
§3.1 熵	38
3.1.1 概念的导出	38
3.1.2 概念与性质	41
3.1.3 熵与通信理论	46
§3.2 Kullback-Leibler 散度	49
3.2.1 定义	49
3.2.2 两个关于信息的不等式	51
3.2.3 在机器学习中的应用：语言生成模型	52
§3.3 附录：Shannon 定理的证明	53
§3.4 习题	54
§3.5 章末注记	56
<b>第四章 Johnson-Lindenstrauss 引理</b>	<b>58</b>
§4.1 机器学习中的数据	58
§4.2 矩法与集中不等式	59
§4.3 J-L 引理的陈述与证明	63
§4.4 J-L 引理的应用	67
§4.5 习题	68
§4.6 章末注记	68
<b>第五章 差分隐私</b>	<b>69</b>
§5.1 数据隐私问题	69
§5.2 差分隐私的定义与性质	71
§5.3 差分隐私的应用	75
5.3.1 随机反应算法	75
5.3.2 全局灵敏度与 Laplace 机制	76
5.3.3 DP 版本 Llyod 算法	78
§5.4 差分隐私与信息论	79
§5.5 习题	80
§5.6 章末注记	80

第三部分 决策与优化	81
第六章 凸分析	82
§6.1 决策与优化的基本原理	82
6.1.1 统计决策理论	82
6.1.2 优化问题	83
6.1.3 例子：网格搜索算法	86
§6.2 凸函数	88
§6.3 凸集	91
6.3.1 基本定义和性质	91
6.3.2 分离超平面定理	93
第七章 对偶理论	95
§7.1 条件极值与 Lagrange 乘子法	96
§7.2 Karush–Kuhn–Tucker 条件	99
§7.3 Lagrange 对偶	102
7.3.1 Lagrange 定理	102
7.3.2 弱对偶定理，强对偶定理	106
§7.4 应用：支持向量机 (SVM)	110
第八章 不动点理论	113
§8.1 Banach 不动点定理	113
§8.2 Brouwer 不动点定理	116
§8.3 不动点的一般视角	119
第四部分 逻辑与博弈	120
第九章 动态博弈	121
§9.1 输赢博弈	121
§9.2 随机博弈 (Markov 博弈)	126
第十章 静态博弈	132
§10.1 正则形式博弈	132
10.1.1 生成对抗网络	133
10.1.2 混合策略	135

§10.2 不完全信息博弈 (Bayes 博弈)	136
<b>第五部分 认知逻辑</b>	<b>141</b>
<b>第十一章 模态逻辑基础</b>	<b>142</b>
§11.1 模态逻辑的起源	142
11.1.1 三段论	142
11.1.2 非经典逻辑	143
§11.2 模态语言	144
§11.3 Kripke 语义与框架语义	147
§11.4 模态可定义性	152
<b>第十二章 认知逻辑与共同知识</b>	<b>154</b>
§12.1 “泥泞的儿童”谜题	154
§12.2 认知逻辑的基本模型与性质	156
12.2.1 “泥泞的儿童”再回顾	160
12.2.2 Aumann 结构	161
§12.3 对不一致达成一致	162
§12.4 Rubinstein 电子邮件博弈	165
<b>第六部分 附录：预备知识</b>	<b>169</b>
<b>附录 A 线性代数基础</b>	<b>170</b>
§A.1 线性空间	170
§A.2 线性映射	174
§A.3 矩阵	179
§A.4 双线性型与二次型	185
§A.5 带内积的线性空间	189
§A.6 行列式	195
§A.7 算子范数与谱理论	198
<b>附录 B 微分学基础</b>	<b>204</b>
§B.1 点集拓扑	204
B.1.1 度量空间, 范数	204

B.1.2	开集与闭集	207
B.1.3	紧致性, 收敛性, 完备性	210
B.1.4	连续映射	213
B.1.5	与实数序有关的性质	216
§B.2	一元函数的微分学	218
B.2.1	导数与微分的定义	219
B.2.2	微分学基本定理	222
§B.3	多元函数的微分学	224
B.3.1	微分、偏导数与导数的定义	224
B.3.2	微分学基本定理	230
B.3.3	隐函数定理	232
附录 C	概率论基础	236
§C.1	从朴素概率论到公理化概率论	236
C.1.1	Kolmogorov 概率论	236
C.1.2	条件概率, 独立性	240
§C.2	随机变量, 分布函数	244
C.2.1	基本定义	244
C.2.2	离散型随机变量	248
C.2.3	连续型随机变量	248
C.2.4	随机向量, 条件分布, 独立性	252
C.2.5	随机变量(向量)的函数	256
§C.3	随机变量的数字特征, 条件数学期望	259
C.3.1	数学期望, Lebesgue 积分	259
C.3.2	数学期望的性质	263
C.3.3	随机变量的内积空间	266
C.3.4	特征函数	268
C.3.5	条件数学期望	269
§C.4	多元正态分布 (Gauss 向量)	273

# 第一部分

## AI 的逻辑



# 第一章 合情推理

1983 年，心理学家 Amos Tversky 和 Daniel Kahneman 进行了一个著名的实验，他们向一组被试提出了如下问题：Linda 是一位单身、外向且年龄为 31 岁的女性。在大学期间，她主修哲学，十分关注种族歧视和社会公正问题，而且曾参加过反核游行。请给下面几个选项发生的可能性排序，1 代表最可能，8 代表最不可能：

1. Linda 是一个小学老师。
2. Linda 在书店工作，并且参加瑜伽课程。
3. Linda 非常积极参加女权运动。
4. Linda 是一个精神病的护理员。
5. Linda 是美国女性选民联盟的成员。
6. Linda 是一名银行出纳员。
7. Linda 是一个保险销售员。
8. Linda 是一名银行出纳员，同时她还非常积极参加女权运动。

他们对以下三组人进行了同样的实验：(1) 88 名来自斯坦福大学或不列颠哥伦比亚大学（UBC）的本科生，没有学过概率统计，(2) 53 名斯坦福大学的一年级研究生，他们学过一门或更多的概率统计课程，(3) 32 名斯坦福大学商学院博士生，他们学过很多概率统计和决策理论的进阶课程。

我们重点关注第三、六、八个选项。抽象地说，我们可以把这三个陈述分别记为  $A$ ,  $B$ ,  $A \wedge B$ 。从逻辑的角度来看， $A$  或者  $B$  是  $A \wedge B$  的子集，因此更可能发生。然而，让人大跌眼镜的是，实验结果显示，所有三组人的打分中，超过 85% 的人认为  $A \wedge B$  更可能发生！

以上现象被称为合取谬误, 它表明人类的推理并不等同于我们所理解的逻辑推理(被称为演绎推理). 合取谬误的背后隐藏了一种新的推理方式, 被称为合情推理. 演绎推理很早就有了非常严格的逻辑学框架, 得到了充分的研究. 然而, 合情推理同样甚至更加重要, 它不仅是人类与生俱来的能力, 更是科学研究和 AI 的逻辑基础.

本章试图建立合情推理的一个数学模型, 这一模型与 Bayes 概率论紧密地联系在一起. 我们会用它来研究一些有趣的推理谬误, 包括合取谬误, 并与演绎推理进行对比. 最后, 我们会讨论 Bayes 概率论中先验为什么是存在的.

## §1.1 命题逻辑的演绎推理

作为回顾, 本节讨论演绎推理, 它最主要的应用场景是数学证明和一些逻辑论证. 演绎推理定义了数学中一些非常重要的概念, 例如“公理”“定理”“证明”. 我们会看到, 与演绎推理密切相关的是形式推理系统, 它尝试将推理的过程抽象为一堆字符串的变换. 作为介绍, 我们这里给出命题逻辑的形式推理系统, 它是最基本的形式推理系统.

命题是一种陈述, 它可以被判定为真或者假. 例如, “北京是中国的首都.” 是一个命题, 而“北京是中国的首都吗?” 是一个疑问句, 所以不是命题. 然而, 利用自然语言定义命题, 很难判断一些奇怪的句子是不是命题:

**例 1.1** 考虑句子  $A$ : “这句话是假的.” 是一个命题吗? 首先,  $A$  是一个陈述句. 那它的真假呢?

- 如果  $A$  是真的,  $A$  说的是“这句话是假的”, 这与  $A$  真矛盾, 所以它不是真的.
- 如果  $A$  是假的,  $A$  的含义是“这句话是真的”, 这与  $A$  假矛盾, 所以它不是假的.

因此, 这句话既不是真的也不是假的!

以上的现象被称为自指, 也就是一个东西在描述自己.

自指为讨论命题的概念带来了非常大的麻烦. 为了避免这些麻烦, 我们回避命题到底是什么这一哲学问题, 而是把他们抽象为一些字符串, 这就是命题公式:

**定义 1.1 (命题公式)** 命题公式 (或简称公式) 是由一些命题变元和连接词组成的字符串, 它们满足以下递归定义:

- 命题变元  $p, q, r, \dots$  是命题公式.
- 如果  $\phi$  和  $\psi$  是命题公式, 那么  $(\neg\phi)$ ,  $(\phi \vee \psi)$ ,  $(\phi \wedge \psi)$ ,  $(\phi \leftrightarrow \psi)$  和  $(\phi \rightarrow \psi)$  都是命题公式.

$\neg, \vee, \wedge, \leftrightarrow, \rightarrow$  被称为连接词,  $\neg$  被称为否定,  $\vee$  被称为析取 (或),  $\wedge$  被称为合取 (与),  $\leftrightarrow$  被称为双向蕴含 (等价),  $\rightarrow$  被称为蕴含 (推出). 在不产生混淆的时候也会省略括号.  $A \wedge B$  也常写作  $AB$ ,  $\neg A$  也常写作  $\bar{A}$ .

**例 1.2** ( $p \vee (q \rightarrow r)$ ) 是命题公式, 但是  $p \vee \vee q$  不是.

命题最重要的概念是“真假”, 因而命题逻辑一个重要的问题是: 什么样的公式是真的? 给定一个公式集  $\Gamma$ , 对一个公式  $\phi$ , 我们有两种真的概念:

- 语义:  $\Gamma \models \phi$ .
- 语形:  $\Gamma \vdash \phi$ .

首先看语义. 所谓语义, 就是“句子的含义”. 在命题逻辑中, 句子就是命题公式, 而含义就是真假. 既然命题公式是递归定义的, 命题公式的真假我们也应该递归定义.

**定义 1.2 (赋值)** 给定命题变元集合  $V$ , 它形成的命题公式的集合是  $L$ , 一个赋值是一个从  $L$  到  $\{\top, \perp\}$  的映射  $v$ , 递归定义如下:

- 如果  $\phi = p \in V$ , 那么  $v(\phi) = v_p$ .
- 如果  $\phi = \neg\psi$ , 那么  $v(\phi) = \begin{cases} \top, & v(\psi) = \perp, \\ \perp, & v(\psi) = \top. \end{cases}$
- 如果  $\phi = \psi \vee \chi$ , 那么  $v(\phi) = \begin{cases} \top, & v(\psi) = \top \text{ 或 } v(\chi) = \top, \\ \perp, & v(\psi) = \perp \text{ 且 } v(\chi) = \perp. \end{cases}$
- 如果  $\phi = \psi \wedge \chi$ , 那么  $v(\phi) = \begin{cases} \top, & v(\psi) = \top \text{ 且 } v(\chi) = \top \\ \perp, & v(\psi) = \perp \text{ 或 } v(\chi) = \perp. \end{cases}$
- 如果  $\phi = \psi \rightarrow \chi$ , 那么  $v(\phi) = \begin{cases} \top, & v(\psi) = \perp \text{ 或 } v(\chi) = \top, \\ \perp, & v(\psi) = \top \text{ 且 } v(\chi) = \perp. \end{cases}$
- 如果  $\phi = \psi \leftrightarrow \chi$ , 那么  $v(\phi) = \begin{cases} \top, & v(\psi) = v(\chi) \\ \perp, & v(\psi) \neq v(\chi) \end{cases}$

以上定义看起来比较冗长, 但是符合我们的直觉. 一种更为直观的表达方式是真值表. 注意, 连接词可以被理解为一个真值向量到真值的映射, 例如  $\wedge$  可以被理解为  $\{\top, \perp\}^2 \rightarrow \{\top, \perp\}$  的映射, 这一映射可以用真值表表示:

$p$	$q$	$p \wedge q$
$\top$	$\top$	$\top$
$\top$	$\perp$	$\perp$
$\perp$	$\top$	$\perp$
$\perp$	$\perp$	$\perp$

于是，定义中的递归可以用直接用  $\wedge$  来表示：如果  $\phi = \psi \wedge \chi$ ，那么  $v(\phi) = v(\psi) \wedge v(\chi)$ 。  
我们看一个例子。

**例 1.3** 考虑公式  $(p \wedge q) \rightarrow r$ ，我们来定义赋值  $v$ 。首先给命题变元赋值：

	$p$	$q$	$r$
$v$	$\perp$	$\top$	$\perp$

然后递归计算公式的真值。首先计算  $(p \wedge q)$  的真值。根据真值表，我们有

$$v(p \wedge q) = v(p) \wedge v(q) = \perp \wedge \top = \perp.$$

然后计算  $(p \wedge q) \rightarrow r$  的真值：

$$v((p \wedge q) \rightarrow r) = v(p \wedge q) \rightarrow v(r) = \perp \rightarrow \perp = \top.$$

接下来，我们从赋值到推理。推理是从前提出发，得到结论的过程。和赋值相关的推理概念有如下定义：

**定义 1.3 (语义蕴含)** 给定公式集  $\Gamma$  和公式  $\phi$ ，如果对任意赋值  $v$ ，如果  $v(\psi) = \top$  对任意  $\psi \in \Gamma$  都成立，那么  $v(\phi) = \top$ ，则称  $\Gamma$  语义蕴含  $\phi$ ，记作  $\Gamma \models \phi$ 。

换言之，给定一个真的前提  $\Gamma$ ，我们所关注的公式  $\phi$  也是真的。这一概念只关心真值之间的关系，但是至于  $\Gamma$  如何导致  $\phi$  为真，不是这一概念可以刻画的。

**例 1.4** 如果  $\Gamma = \{p \rightarrow q, q \rightarrow r\}$ ，那么  $\Gamma \models p \rightarrow r$ 。我们来证明这个，对任意赋值  $v$ ，如果  $v(p \rightarrow q) = \top$ ，那么  $v(p) = \perp$  或者  $v(q) = \top$ ，

- 如果  $v(p) = \perp$ ，那么  $v(p \rightarrow r) = \top$ 。
- 如果  $v(q) = \top$ ，那么  $v(q \rightarrow r) = v(r)$ 。因为  $v(q \rightarrow r) = \top$ ，所以根据定义  $v(r) = \top$ ，于是  $v(p \rightarrow r) = \top$ 。

不论在何种情况，都有  $v(p \rightarrow r) = \top$ ，即  $\Gamma \models p \rightarrow r$ 。

接下来，我们看语形。语形就是“句子的形态”，或者说一个句子是如何构造出来的。它更像我们所理解的“推理”，也就是从一些公式出发，得出另一些公式。推导规则描述了从一些公式出发如何得到另外公式。语法推导的形式是

$$\frac{\phi_1 \quad \phi_2 \quad \dots \quad \phi_n}{\psi}$$

横线上方的称为前提，横线下方的称为结论。

在一些命题逻辑的形式推理系统（比如 Hilbert 推理系统）中，推导规则只有一条，即肯定前件（MP）：

$$\frac{\phi \rightarrow \psi \quad \phi}{\psi}$$

然而，引入更多推导规则可以有助于简化推理的过程，这是另外一些形式推理系统（比如自然演绎系统）的做法。比如，我们可以加入合取和析取的推导规则：

- 引入新的连接词，例如

$$\frac{\phi \quad \psi}{\phi \wedge \psi}$$

- 消除连接词，例如

$$\frac{\phi \wedge \psi}{\phi}$$

我们再看一个非常重要的例子。

**例 1.5** 将归谬法（RAA）加入推导法则：将  $\neg\phi$  作为前提，得到了结论  $\perp$ ，那么可以推出  $\phi$  才是结论。写作

$$\frac{\begin{array}{c} [\neg\phi] \\ \vdots \\ \perp \end{array}}{\phi}$$

方括号表示假设  $[\neg\phi]$  是前提，省略号表示推导的中间步骤。注意，这里是假设了一个前提，而不是真的前提，这一点是非常重要的。利用这一推导规则，我们就可以得到反证法。

演绎推理所对应的数学模型即为形式推理系统：一系列字符串的集合（命题公式），以及一系列的推导规则，以及公理的集合（一些命题公式）。

在形式推理系统中，我们可以形式地定义什么是“推理”和“结论”，而这些东西构成了演绎推理的过程与结果：

**定义 1.4 (演绎推理)** 设  $L$  是公式集， $A \subseteq L$  是公理的集合， $\mathcal{R}$  是推导规则， $\Gamma \subseteq L$ ，考虑  $L$  中的序列  $\phi_1, \dots, \phi_n$ ，如果对任意  $i$ ， $\phi_i$  满足下面的其中一个条件：

1.  $\phi_i \in A$ ,
2.  $\phi_i \in \Gamma$ , 或者
3. 存在一个推导规则，和  $\phi_{i_1}, \dots, \phi_{i_k}$ ,  $i_1, \dots, i_k < i$  使得  $\phi_i$  可以由  $\phi_{i_1}, \dots, \phi_{i_k}$  推导出来，

那么称  $\phi_1, \dots, \phi_n$  是从  $\Gamma$  出发的一个**演绎推理**，记作  $\Gamma \vdash \phi_n$ ， $\phi_n$  称为**结论**， $\Gamma$  称为**前提**。

简而言之，记号  $\Gamma \vdash \phi$  的意思是以  $\Gamma$  作为前提，依据推导法则，通过一系列（但有限）的推理步骤，可以得到  $\phi$  作为结论。因而，语形定义的真值是“可以被推理出来”的概念。对比语义，语形更加关注“过程”，而不是“关系”。

介绍完了语义和语形，我们现在来看他们之间的关系，其中最重要的是完备性定理：

**定理 1.1 (完备性定理)** 对任意公式集  $\Gamma$  和任意公式  $\phi$ ，

$$\Gamma \models \phi \iff \Gamma \vdash \phi.$$

完备性定理有两层意思：能推理出来的都是真的，真的都能被推理出来。并非所有的形式推理系统都有完备性定理，但是对于命题逻辑来说，这个定理成立的形式推理系统是更加符合直观的。

完备性定理的一个直接推论是，我们可以用真值表来检查从前提出发是否能推出结论。例如，在例 1.4 中我们验证了  $p \rightarrow q, q \rightarrow r \vdash p \rightarrow r$ <sup>1</sup>，所以也有

$$p \rightarrow q, q \rightarrow r \vdash p \rightarrow r,$$

即便我们完全不知道这一推理是如何进行的，我们依然知道这一推理是可行的！

如果  $\emptyset \vdash \phi$ ，那么称  $\phi$  为重言式。重言式是不需要加任何假设也一定成立的公式，是这一推理系统所包含的“正确的废话”。如果  $\psi \leftrightarrow \phi$  是重言式，我们就说  $\psi, \phi$  是等值的。等值的公式在演绎推理中可以互相替代使用。例如： $p \rightarrow q$  与  $\neg q \rightarrow \neg p$  是等值的，所以证明一个命题也可以去证明它的逆否命题。

---

<sup>1</sup>在书写时，我们常常省略  $\Gamma$  中的花括号，因此这里写作  $p \rightarrow q, q \rightarrow r \vdash p \rightarrow r$  而不是  $\{p \rightarrow q, q \rightarrow r\} \vdash p \rightarrow r$ 。对于符号  $\vdash$  我们遵循同样的规则。

## §1.2 合情推理的数学模型

现在，我们将话题转到合情推理。出乎意料的是，人类先天所具备的推理能力，更接近合情推理而不是演绎推理。在婴儿时期，我们需要建立各式各样的经验规则，以便在这个世界生存下来。例如，如果婴儿碰到了带电的插座，那么他/她会感受到疼痛，这样就知道了插座是危险的。下次遇到插座的时候，他/她就会避开插座，因为他/她学到了“所有插座都是危险的”这一经验。上面的例子都是从一些特殊的事件（摸插座触电了）出发，得到了一些一般的结论（所有插座都是危险的），这就是合情推理的基本形式。

为了从更抽象的角度理解合情推理的形式，我们用经典的三段论来对比演绎推理和合情推理。

演绎推理包含两个强三段论：

$$\begin{array}{cc} \phi \rightarrow \psi & \phi \rightarrow \psi \\ \phi \text{ 是真的} & \psi \text{ 是假的} \\ \hline \psi \text{ 是真的} & \phi \text{ 是假的} \end{array}$$

例如，从科学的角度，触电会疼痛 ( $\phi \rightarrow \psi$ )，摸插座会触电 ( $\phi$ )，所以摸插座会疼痛 ( $\psi$ )。这是一个演绎推理的例子。

然而，合情推理是完全发过来的，即弱三段论，形式上写作

$$\begin{array}{c} \phi \rightarrow \psi \\ \psi \text{ 是真的} \\ \hline \phi \text{ 变得更合理} \end{array}$$

例如，触电会疼痛 ( $\phi \rightarrow \psi$ )，摸插座会疼痛 ( $\psi$ )，所以摸插座很可能会触电 ( $\phi$ )。这样的推理更符合上面婴儿推理的模式。虽然我们不能得知  $\phi$  的真假，但是根据已有的事实  $\psi$ ，我们可以得到  $\phi$  的合理性推断。这样没有严格真假的推理模式就是合情推理。

推理的一个重要概念是“真假”，或者说这个推理的正确性、合理性。在演绎推理中，推理只有两种可能：正确的推理和错误的推理；而在合情推理中，推理的合理性有多种可能，例如“很可能正确”“可能正确”“很可能错误”等。我们来看一个例子。

**例 1.6** 假如婴儿摸了插座，并且感到了疼痛，那么，你觉得他/她会怎么想？

- 选项 1：摸插座是危险的。
- 选项 2：摸的插座上面有微小不可见的细针，细针扎了他/她的手指。

虽然选项 1 和选项 2 都是可能的，但是他们的合理程度并不同：选项 1 明显比选项 2 要合理得多。这里产生了一个自然的问题：如何刻画不同推理的合理程度？

### 1.2.1 合情推理的基本假设，似然

接下来，我们给合情推理建立一个数学模型。合情推理的基本方式是给定已知的命题  $B$ ，推出命题  $A$  的合理程度，比如“非常可能是真的”“可能是真的”“可能是假的”“非常可能是假的”等，我们用符号  $A|B$  表示这一合理程度。

在写出符号的时候，我们已经假设了合理程度所对应的数学对象： $A|B$  应该是  $L \times L \rightarrow \mathbb{R}$  的函数，这里  $L$  是命题公式的集合。但是，这一表述是有瑕疵的，因为不是所有的前提到结论的推理都可以用合理程度来描述。比如， $B$  是“太阳从西边升起”，前提根本就是一个假命题，从假命题出发，何谈“合情”？所以，符号  $A|B$  应该只对一些特定的  $A, B$  有意义，我们称之为  $L \times L \rightarrow \mathbb{R}$  的偏函数，并将这一函数写为  $f(A|B)$ ，称作似然。

我们还没有确定这一映射的具体形式。但是，这样的映射要满足两个原则：1. 符合我们对于合情推理的基本直观，2. 易于计算。

在这里，我们类比行星运动定律来理解这第二个原则的具体含义。最开始的时候，人们认为天体都是绕着地球转的。基于这样的思想，Ptolemaeus 提出了行星运动的模型。在这个模型中，Ptolemaeus 引入了“均轮”和“本轮”的概念。行星在一个较小的圆（本轮）上做匀速圆周运动，而这个本轮的中心又在一个更大的圆（均轮）上绕地球做匀速圆周运动。这一模型可以较为准确预测天体的运动，然而需要非常复杂的概念。

后来，Kepler 提出了完全不同的行星运动模型：行星绕着恒星做椭圆运动，与恒星的连线扫过的面积与运动时间成正比，行星轨道的半长轴的三次方与其公转周期的二次方的比值都相等。这就是著名的“Kepler 三定律”。在 Kepler 的模型中，我们放弃了“均轮”和“本轮”的概念，而是用简单的椭圆来描述行星的运动。尽管 Ptolemaeus 模型和 Kepler 模型都可以描述行星的运动，但是 Kepler 模型更加简单，更加容易计算，也正是在此基础上，Newton 才发现了万有引力定律。

同样的道理也适用于合情推理。或许会有很多的  $f$  可以符合直观地描述合理程度，但是我们希望找到一个尽可能容易使用和计算的  $f$ ，它能有更广泛的应用，或许也蕴含了更深层次的规律。现在，我们暂时将所有符合合情推理直观的映射的集合记作  $\mathcal{L}$ 。我们给出  $\mathcal{L}$  中的函数要具备的性质（即假设），然后再确定具体的函数。

首先，在命题逻辑中，我们只关心命题的真假，并不关心命题具体是什么。在合情推理中，我们依然希望有这样的性质：推理的合理程度不依赖具体的命题形式。在数学上，这等价于如下的原则：

**规则 1.1 (等值原则)** 对任意  $f \in \mathcal{L}$ ，任意命题公式  $\phi_1, \phi_2, \psi \in L$ ，如果  $\phi_1$  和  $\phi_2$  等值（即  $\phi_1 \leftrightarrow \phi_2$  是重言式），那么

$$f(\phi_1|\psi) = f(\phi_2|\psi).$$



例如,  $\phi \rightarrow \psi$  和  $\neg\phi \vee \psi$  的具体含义是不同的: 一个是蕴含式 (即  $\phi$  推出  $\psi$ ), 一个是两个命题 ( $\neg\phi$  和  $\psi$ ) 的或; 但是他们等值, 所以  $f(\phi \rightarrow \psi|\xi) = f(\neg\phi \vee \psi|\xi)$ . 更多的例子如下:

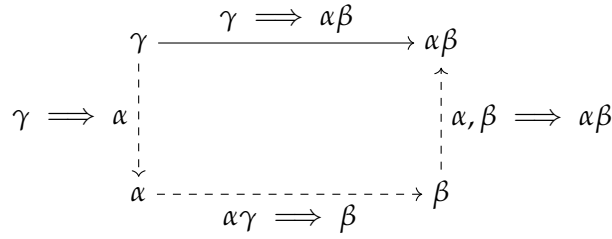
**例 1.7** 对任意  $f \in \mathcal{L}$  以及公式  $\alpha, \beta, \gamma$ , 成立

- $f(\alpha \vee \beta|\gamma) = f((\beta \vee \alpha)|\gamma)$ .
- $f((\alpha\beta)\gamma|\delta) = f(\alpha(\beta\gamma)|\delta)$ .
- $f(\neg\neg\alpha|\beta) = f(\alpha|\beta)$ .

现在来看第二条假设, 它与推理不同的路径有关。考虑从  $\gamma$  出发推出  $\alpha\beta$  为真。我们有两种推理的链条:

1. 直接推理: 直接从  $\gamma$  推出  $\alpha\beta$ .
2. 分步推理: 先从  $\gamma$  推出  $\alpha$ , 然后从  $\alpha\gamma$  推出  $\beta$ , 因此自然也推出了  $\alpha\beta$ .

也可以用下面的图来直观表示, 其中实线箭头表示直接推理, 虚线箭头表示分步推理:



我们前面要求, “推理的合理程度不依赖具体的命题形式”, 现在, 我们将这一要求进一步细化为“推理的合理程度不依赖推理的具体形式”. 因为上面的直接推理和分步推理都是从  $\gamma$  推出  $\alpha\beta$ , 所以我们要求他们的合理程度相同, 即

**假设 1.1** 对任意  $f \in \mathcal{L}$ , 存在一个  $\mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  的函数  $F$ , 对任意命题公式  $\alpha, \beta, \gamma$ ,

$$f(\alpha\beta|\gamma) = F(f(\alpha|\gamma), f(\beta|\alpha\gamma)).$$

换言之, 在考虑我们不关心  $\gamma$  具体是怎么推出来  $\alpha$  的, 我们只要求  $\alpha\beta$  的合理程度与  $\alpha$  和  $\beta$  的合理程度有关.

注意, 以上假设不能写成  $f(\alpha\beta|\gamma) = F(f(\beta|\gamma), f(\alpha|\gamma))$ . 例如, 假设  $\alpha$  是“小明的左眼是蓝色的”,  $\beta$  是“小明的右眼是棕色的”.  $f(\alpha|\gamma)$  和  $f(\beta|\gamma)$  都可以很合理, 但  $f(\alpha\beta|\gamma)$  则非常罕见. 同样的道理, 我们可以排除其他形式的假设, 比如  $f(\alpha\beta|\gamma) = F(f(\gamma|\alpha), f(\beta|\gamma))$ .

我们定义的  $F$  具有如下性质:

**定理 1.2** 如果  $F$  是二阶连续可微函数, 那么  $F$  充分必要地满足

$$cf(F(p, q)) = f(p)f(q),$$

这里,  $c$  是某个常数,  $f$  是某个函数. 因此, 对任意命题公式  $\alpha, \beta, \gamma$ , 成立

$$cf(\alpha\beta|\gamma) = f(\alpha|\gamma)f(\beta|\alpha\gamma).$$

回忆, 我们的目标是找到一个尽可能简单的  $f$ , 它能够描述合情推理的合理程度. 因此, 我们可以接受  $F$  二阶连续可微的假设, 然后让  $c = 1$ . 于是, 我们得到了合情推理的一个基本原则:

**规则 1.2 (与规则)** 对命题  $\alpha, \beta, \gamma$ , 成立

$$f(\alpha\beta|\gamma) = f(\alpha|\gamma)f(\beta|\alpha\gamma).$$

第二个考虑的是似然  $f(\alpha|\beta)$  和  $f(\neg\alpha|\beta)$  的关系. 两个似然应该存在某种函数关系, 于是我们有如下假设:

**假设 1.2** 对任意  $f \in \mathcal{L}$ , 存在函数  $S: \mathbb{R} \rightarrow \mathbb{R}$ , 对任意命题公式  $\alpha, \beta$ , 成立

$$f(\neg\alpha|\beta) = S(f(\alpha|\beta)).$$

在与规则限制之下, 我们可以证明  $S$  具有如下性质:

**定理 1.3** 如果  $S$  二阶连续可微, 那么  $S$  充分必要地满足

$$S(x) = (1 - x^m)^{1/m},$$

这里的  $m$  是一个正常数.

由上面的定理, 我们可以得到以下性质

**推论 1.1**  $f(\alpha|\beta)^m + f(\neg\alpha|\beta)^m = 1$ .

**证明.**  $f(\neg\alpha|\beta) = S(f(\alpha|\beta)) = (1 - f(\alpha|\beta)^m)^{1/m}$ . □

同样, 从简洁性的角度出发, 我们可以接受  $S$  二阶连续可微的假设, 然后让  $m = 1$ . 于是, 我们得到了合情推理的另一个基本原则:

**规则 1.3 (否定规则)** 对命题  $\alpha, \beta$ , 成立

$$f(\alpha|\beta) + f(\neg\alpha|\beta) = 1.$$

由上，我们得到了合情推理的三个基本原则：等值原则、与规则和否定规则。这三个原则构成了合情推理的基本框架。

接下来，我们说明，这三个原则就已经足够用来计算任意的合情推理。首先，连接词  $\{\wedge, \neg\}$  组成了演绎逻辑的完备集：所有真值函数都可以用这个集合的词来表示，因此，根据等值规则，只要可以计算包含  $\wedge$  和  $\neg$  的合情推理，就可以计算任意合情推理。其次，我们可以用与规则和否定规则来计算  $\wedge$ ，否定规则计算  $\neg$ ，因此，任何命题都可以通过这三个原则来计算。

**例 1.8** 例如，计算  $f(A \vee B|C)$ ，

$$\begin{aligned}
 f(A \vee B|C) &= 1 - f(\overline{A} \wedge \overline{B}|C) \\
 &= 1 - f(\overline{A}|C)f(\overline{B}|\overline{A}C) \\
 &= 1 - f(\overline{A}|C)(1 - f(B|\overline{A}C)) \\
 &= f(A|C) + f(\overline{A}B|C) \\
 &= f(A|C) + f(B|C)f(\overline{A}|BC) \\
 &= f(A|C) + f(B|C)(1 - f(A|BC)) \\
 &= f(A|C) + f(B|C) - f(AB|C).
 \end{aligned}$$

### 1.2.2 似然与概率

我们已经知道，似然是一个合理性度量，在本节中，我们将似然与概率联系起来：他们实际上是等价的。

首先介绍 *Kolmogorov* 的概率论。这一概率论研究事件空间上的概率测度。其研究对象的总体被称为样本空间，记为  $\Omega$ 。我们只能观测到某种可观测的特性  $P$ ，而不能直接观测样本点，即我们只能观察事件，或者说集合

$$\{\omega \in \Omega : P(\omega)\}.$$

我们可以观测的所有事件的集合称为事件域，记为  $\mathcal{F}$ 。

事件域  $\mathcal{F}$  中的事件之间互相有关联。我们自然可以观测到  $\Omega$ ，因此  $\Omega \in \mathcal{F}$ 。如果我们可以观测到事件  $A$ ，那么我们也可以通过没有观测到  $A$  来判断观测到了  $\Omega \setminus A$ 。因此， $A \in \mathcal{F} \implies \Omega \setminus A \in \mathcal{F}$ 。如果我们观测到了  $A$  或者  $B$ ，我们其实也观测到了  $A \cup B$ ，即  $A, B \in \mathcal{F} \implies A \cup B \in \mathcal{F}$ 。

事实上，我们可以形式上定义概率：

**定义 1.5 (概率)** 概率  $\text{Pr}$  是一个事件域  $\mathcal{F}$  到实数的映射，并且满足：

- 规范性:  $\Pr(\Omega) = 1$ .
- 非负性: 对任意  $A \in \mathcal{F}$ ,  $\Pr(A) \in [0, 1]$ .
- 可列可加性: 对  $A_1, A_2, \dots$  满足  $A_i \cap A_j = \emptyset, i \neq j$ , 有

$$\Pr\left(\bigcup_i A_i\right) = \sum_i \Pr(A_i).$$

事件是命题的集合论描述. 具体来说, 有如下对应

事件	命题
$\Omega$	$\top$
$\emptyset$	$\perp$
$A \cap B$	$A \wedge B$
$A \cup B$	$A \vee B$
$A \subseteq B$	$A \rightarrow B$
$A = B$	$A \leftrightarrow B$

回忆条件概率的定义, 我们可以得到链式法则:

$$\Pr(AB|C) = \frac{\Pr(ABC)}{\Pr(C)} = \frac{\Pr(B|AC) \Pr(AC)}{\Pr(C)} = \Pr(B|AC) \Pr(A|C).$$

回忆补事件公式:  $\Pr(A|C) + \Pr(\bar{A}|C) = \Pr(\Omega|C) = 1$ . 这两个公式恰好对应了合情推理的与规则以及否定规则!

这并不是巧合, 实际上, 合情推理与事件域的公理化概率具有一一对应的关系, 见表 1.1.

公理化概率论	合情推理
事件	命题
(条件) 概率	似然
链式法则	与规则
补事件公式	否定规则

表 1.1: 概率论与合情推理的对应

从这个意义上说, 概率是似然唯一的数学模型! 从此, 我们将似然  $f(A|B)$  定义为概率  $\Pr(A|B)$ .

### 1.2.3 先验与基率谬误

在前面，我们有意模糊了条件概率和概率在合情推理中的区别。然而，这样的区别是非常重要的。在合情推理中，非条件的概率被称为先验概率或者基率，它表示了对这个命题合理程度的一种无条件的信念。对应地，条件概率就是后验概率或似然，它表示了对合情推理合理程度的一种信念。先验概率和后验概率有若干相互转化的公式。

条件概率有全概率公式：

**定理 1.4 (全概率公式)** 设  $A_i$  是彼此互斥的事件， $\cup_i A_i = \Omega$ ，那么

$$\Pr(B) = \sum_i \Pr(B|A_i) \Pr(A_i).$$

全概率公式表明了如何使用似然建立起不同先验概率之间的联系。

条件概率还有 Bayes 定理：

**定理 1.5 (Bayes 定理)**

$$\Pr(A|B) = \Pr(B|A) \frac{\Pr(A)}{\Pr(B)}.$$

Bayes 定理表明了两个不同的后验概率如何基于先验概率相互转化。在合情推理中，这表明了前提推结果的强三段论和结果推前提合理性的弱三段论之间的关系。

下面我们看一个合情推理的例子。一辆出租车在夜间发生了一起肇事逃逸事故。这座城市有两家出租车公司，绿色和蓝色。这个城市历史上肇事逃逸车辆 85% 是绿色的，15% 是蓝色的。一名目击者指认出租车是蓝色的，这里的指认并不一定正确。考虑一种理想的假设，法庭知道这位证人 80% 的概率能正确识别颜色，20% 的概率会把颜色识别错。问：事故车辆是那种颜色的可能性更大？忽略先验概率会产生答案是蓝车的结论。

下面我们考虑先验概率再次做计算。记  $B$  为肇事逃逸的出租车为蓝色， $G$  为肇事逃逸的出租车为绿色， $R$  为目击者指认出租车是蓝色。先验概率为  $\Pr(B) = 0.15$ ， $\Pr(G) = 0.85$ 。似然为  $\Pr(R|B) = 0.8$ ， $\Pr(R|G) = 0.2$ 。利用全概率公式， $\Pr(R) = \Pr(B) \Pr(R|B) + \Pr(G) \Pr(R|G) = 0.29$ 。

利用 Bayes 定理， $\Pr(B|R) = \Pr(R|B) \Pr(B) / \Pr(R) \approx 0.41$ 。然而， $\Pr(G|R) = \Pr(R|G) \Pr(G) / \Pr(R) \approx 0.59$ 。因此我们更应该倾向于认为肇事逃逸的出租车是绿色的！

上面的例子表明人类决策有一种忽略先验概率的倾向，这种倾向被称为基率谬误。为了研究这种谬误，我们将先验概率也纳入合情推理的模型之中。这样，Kolmogorov 的概率论和合情推理就完全一致了。反过来说，如果我们将概率论理解为如上合情推理的数学模型，我们就得到了 Bayes 概率论。

## §1.3 合情推理的归纳强论证

### 1.3.1 归纳强论证

在基率谬误的例子中，我们看到合情推理必须要完整地考虑先验的影响。另一方面，我们看到最终做出决策的方式是最大似然，即似然更高的那个命题更有可能是对的。这说明合情推理中有很不同于演绎推理的论证方式。

我们首先指出合情推理包含了命题逻辑中的两个强三段论：

$$\begin{array}{cc} A \rightarrow B & A \rightarrow B \\ A \text{ 是真的} & B \text{ 是假的} \\ \hline B \text{ 是真的} & A \text{ 是假的} \end{array}$$

我们以第一个三段论为例，记  $C \equiv A \rightarrow B$ 。由链式法则， $\Pr(B|AC) = \Pr(AB|C) / \Pr(A|C)$ 。  $A \rightarrow B$  意味着作为事件  $A \subseteq B$ ，即  $AB = A$ ， $\Pr(AB|C) = \Pr(A|C)$ 。代入上式得到  $\Pr(B|AC) = 1$ ，这就是说，当  $A$  为真时， $B$  也为真。

除了演绎逻辑中的强三段论，合情推理还包含了弱三段论的定量形式：

$$\begin{array}{c} A \rightarrow B \\ B \text{ 是真的} \\ \hline A \text{ 变得更合理} \end{array}$$

$\Pr(A|C)$  是  $A$  的似然，而  $\Pr(A|BC)$  是假设  $B$  为真时， $A$  的似然。由链式法则， $\Pr(A|BC) = \Pr(A|C) \Pr(B|AC) / \Pr(B|C)$ 。因为  $\Pr(B|AC) = 1$  且  $\Pr(B|C) \leq 1$ ，所以  $\Pr(A|BC) \geq \Pr(A|C)$ 。也就是当  $B$  为真时， $A$  的合理程度会变大。

仿照三段论，现在我们将演绎推理中的若干概念推广到合情推理中。

回忆记号  $X \vdash Y$  表示从前提为  $X$  出发可以跟据推导规则得到结论  $Y$ 。此时，我们说从  $X$  到  $Y$  的过程是一个有效论证，或者说  $Y$  是  $X$  的逻辑结论。它与以下三个表述等价：

- $X \vdash Y$ 。
- $X \rightarrow Y$  是重言式。
- $X \wedge \neg Y$  是矛盾式。

等价性证明由蕴含的推导法则（或演绎定理）和完备性定理得出。那么，如何在合情推理中定义类似的概念呢？为此，我们引入随机真值表。

回忆, 事件是命题的集合论描述. 合情推理中, 每个事件被赋予一个概率 (似然), 对应的命题也会被赋予同样的概率. 于是对应于演绎推理中的语义真值表, 合情推理中有随机真值表, 例如

Pr	A	B	$A \vee B$
0.4	⊤	⊤	⊤
0.2	⊤	⊥	⊤
0.25	⊥	⊤	⊤
0.15	⊥	⊥	⊥

在合情推理中, 我们也有和有效论证对应的归纳强论证. 考虑如下推理:

$$\frac{X}{Y}$$

利用随机真值表, 我们可以尝试定义归纳强论证为  $X \wedge \neg Y$  的不太可能为真 (或  $\neg X \vee Y$  很可能为真). 然而我们会看到, 仅仅用随机真值表得到的概念是不符合合情推理的直觉的. 我们通过两类例子来引入归纳强论证的限制条件.

**例 1.9 (奇怪的例子一)** 记  $X$  为一个北京大学的同学今年 2000 岁,  $Y$  为一个北京大学的同学今年 2000 岁, 并且有三条胳膊. 直观来讲, 如上  $X \vdash Y$  不是归纳强论证. 但是, 其等价的公式  $\neg X \vee Y$ , 成立的概率足够大.

考虑另外一个有效论证中的例子. 记  $A_n = \{\text{圆周率 } \pi \text{ 的系数有 } n \text{ 个连续的 } 1\}$ . 那么, 命题  $A_{n+1} \rightarrow A_n$  看起来是显然的, 但与之等值的命题  $\neg A_{n+1} \vee A_n$  却不太容易被一眼接受.

这两个例子都说明, 判断是否为归纳强论证不能只关注结论成立的概率. 于是, 我们得到限制条件: 证据支持, 它的定义如下.

**定义 1.6 (证据支持)** 假设  $X$  和  $Y$  是命题公式,  $t \in [0.5, 1]$ . 如果  $\Pr(Y|X) > t$ , 我们称证据  $X$  支持  $Y$  的强度大于  $t$ .

证据支持是比最大似然准则的更进一步的要求. 然而, 证据支持并不能解决所有问题. 考虑下面的例子.

**例 1.10 (奇怪的例子二)** 记  $X$  为小明是一位北京大学的学生,  $Y$  为小明不会飞. 表面看来,  $\Pr(Y|X) = 1$ , 但我们知道  $X \vdash Y$  并不应该是归纳强论证. 问题出在了  $\Pr(Y)$  本身就等于 1, 所以  $\Pr(Y|X) = 1$  并没有什么实际意义.

从这个例子出发, 我们得到另一个限制条件: 正相关性, 它的定义如下.

**定义 1.7 (相关性)** 我们称  $X$  与  $Y$  正相关, 如果  $\Pr(Y|X) > \Pr(Y)$ . 等价地, 示性函数  $I(X)$  和  $I(Y)$  相关系数大于 0. 类似地, 如果  $\Pr(Y|X) < \Pr(Y)$  (或  $I(X)$  和  $I(Y)$  的相关系数小于 0), 那么  $X$  和  $Y$  负相关. 如果  $\Pr(Y|X) = \Pr(Y)$  (或  $I(X)$  和  $I(Y)$  的相关系数等于 0), 那么  $X$  和  $Y$  不相关. 在归纳强论证中, 我们要求  $X$  和  $Y$  正相关.

加上以上两个限制条件, 我们可以得到归纳强论证的严格定义.

**定义 1.8 (归纳强论证)** 如果论证  $X \Rightarrow Y$  满足以下三个条件, 我们称之为归纳强论证:

- $X$  证据支持  $Y$ :  $\Pr(Y|X) > 0.5$ .
- $X$  与  $Y$  正相关:  $\Pr(Y|X) > \Pr(Y)$ .
- $X \rightarrow Y$  不是有效论证.

不将有效论证定义为归纳强论证得原因之一是: 一个论证可以在前提  $X$  是矛盾式时有效. 例如:  $P \wedge \neg P \vdash Q$  是有效论证. 但是, 由于  $\Pr(P \wedge \neg P) = 0$ ,  $\Pr(Q|P \wedge \neg P)$  是无定义的, 所以  $P \wedge \neg P$  并不证据支持  $Q$ , 也不和  $Q$  正相关.

进一步, 我们还希望能够衡量前提  $X$  在多大程度上确认结论  $Y$  成立, 我们可以通过如下两个量来衡量:

**定义 1.9 (认可度)** 考虑命题  $X$  和  $Y$ . 我们定义认可概率增量和认可度似然比如下:

- 认可概率增量  $d(X, Y)$ :  $X$  发生后给  $Y$  的发生增加了多大的概率.

$$d(X, Y) = \Pr(Y|X) - \Pr(Y).$$

- 认可度似然比  $\ell(X, Y)$ :  $Y$  发生时  $X$  的似然会比  $Y$  没发生时  $X$  的似然增加多少. 该差值越大表示观测到  $X$  的话越应该发生了  $Y$ . 分母归一化使得  $\ell(X, Y) \in [-1, 1]$ .

$$\ell(X, Y) = \frac{\Pr(X|Y) - \Pr(X|\neg Y)}{\Pr(X|Y) + \Pr(X|\neg Y)}.$$

这两个认可度的定义其实相互可以转化, 我们将这一转化留作习题. 利用这一转化, 认可度和相关性的关系可以表达如下:

**命题 1.1** 设  $0 < \Pr(X), \Pr(Y) < 1$ , 下列等价成立:

- $X$  和  $Y$  正相关  $\iff d(X, Y) > 0 \iff \ell(X, Y) > 0$ .
- $X$  和  $Y$  不相关  $\iff d(X, Y) = 0 \iff \ell(X, Y) = 0$ .



- $X$  和  $Y$  负相关  $\iff d(X, Y) < 0 \iff \ell(X, Y) < 0$ .

认可度与有效论证的关系可以表达如下:

**命题 1.2** 设  $0 < \Pr(X), \Pr(Y) < 1$ , 那么

$$d(X, Y) = \begin{cases} \Pr(\neg Y), & \text{如果 } X \vdash Y, \\ -\Pr(Y), & \text{如果 } X \vdash \neg Y. \end{cases}$$

$$\ell(X, Y) = \begin{cases} 1, & \text{如果 } X \vdash Y, \\ -1, & \text{如果 } X \vdash \neg Y. \end{cases}$$

### 1.3.2 有效论证和归纳强论证的比较

考虑一个论证  $X \Rightarrow Y$ , 我们已经有三种方式评估  $X$  如何支持  $Y$ :

1.  $X \Rightarrow Y$  是一个演绎推理:  $X \vdash Y$ .
2. 基于随机真值表,  $X$  证据支持  $Y$ :  $\Pr(Y|X) > 0.5$ .
3. 基于随机真值表,  $X$  正相关于  $Y$ :  $\Pr(Y|X) > \Pr(Y)$ .

其中, 1 对应有效论证, 2 和 3 都是合情推理中的归纳强论证的必要条件. 我们将进一步讨论有效论证和归纳强论证的一些不同之处.

首先, 有效论证具有单调性: 论证的有效性随着前提的增加不会下降. 即: 对于任意  $X, Y, Z$ , 若  $X \vdash Y$ , 则  $X, Z \vdash Y$ . 然而合情推理中, 单调性不再存在. 请看下面的例子.

**例 1.11 (非单调性: 例子)** 考虑命题  $X, Y, Z$  和对应的随机真值表.

Pr	X	Y	Z	$X \wedge Z$
0.1	T	T	T	T
0.2	T	T	⊥	⊥
0.2	T	⊥	T	T
0	T	⊥	⊥	⊥
0.1	⊥	T	T	⊥
0.1	⊥	T	⊥	⊥
0.1	⊥	⊥	T	⊥
0.2	⊥	⊥	⊥	⊥

$\Pr(Y|X) = (0.1 + 0.2) / (0.1 + 0.2 + 0.2 + 0) = 0.6 > 0.5$ , 然而,  $\Pr(Y|X \wedge Z) = (0.1) / (0.1 + 0.2) = 1/3 < 0.5$ . 因此,  $X$  证据支持  $Y$ , 但  $X \wedge Z$  不支持  $Y$ .

接下来，我们用  $Z$  论证  $Y$ ，将  $X$  看作某种附加的条件，我们考虑  $X$  对  $Y$  这一论据的影响。在演绎推理中，若  $Z, X \vdash Y$  和  $Z, \neg X \vdash Y$  都满足，则  $Z \vdash Y$ 。换言之，不论补充论据  $X$  还是  $\neg X$ ，从  $Z$  出发都可以论证出  $Y$ 。

如果类比到合情推理中呢？这就涉及到确凿性原则：如果不论条件在  $X$  还是  $\neg X$ ， $Z$  都是  $Y$  的一个“好的论据”，那么  $Z$  就是  $Y$  的一个“好的论据”。与之相对应的是无条件确凿性原则：如果  $Z \wedge \neg X$  和  $Z \wedge X$  都是  $Y$  的一个“好的论据”，那么  $Z$  就是  $Y$  的一个“好的论据”。“好的论据”可以从证据支持和正相关性两方面考虑。

在任何随机真值表中，如果  $\Pr(Y|Z \wedge X) > 0.5$  且  $\Pr(Y|Z \wedge \neg X) > 0.5$ ，那么根据全概率公式， $\Pr(Y|Z) > 0.5$ 。因此，从证据支持角度，确凿性原则是成立的。而同样的论证也说明，从证据支持角度，无条件确凿性原则也是成立的。

在任何随机真值表中，如果  $\Pr(Y|Z \wedge X) > \Pr(Y)$  且  $\Pr(Y|Z \wedge \neg X) > \Pr(Y)$ ，那么同样根据全概率公式， $\Pr(Y|Z) > \Pr(Y)$ 。因此，从正相关性角度，无条件确凿性原则是成立的。那么，从正相关性角度，确凿性原则成立吗？

**注。** 在讨论正相关的时候，无条件确凿性原则和确凿性原则容易被混淆。在无条件确凿性原则中，我们要求  $Z \wedge X$  和  $Z \wedge \neg X$  都是  $Y$  的好论据，这表现为  $\Pr(Y|Z \wedge X) > \Pr(Y)$  且  $\Pr(Y|Z \wedge \neg X) > \Pr(Y)$ 。而在确凿性原则中，我们要求不论条件在  $X$  还是  $\neg X$ ， $Z$  都是  $Y$  的好论据。这一定义其实引入了两个新的概率，即  $Q^+(\cdot) = \Pr(\cdot|X)$  和  $Q^-(\cdot) = \Pr(\cdot|\neg X)$ 。他们的定义为

$$Q^+(\cdot) = \frac{\Pr(\cdot \wedge X)}{\Pr(X)}, \quad Q^-(\cdot) = \frac{\Pr(\cdot \wedge \neg X)}{\Pr(\neg X)}.$$

于是，确凿性原则中涉及到了三个正相关：在前提中，我们假设了  $Q^+(Y|Z) > Q^+(Y)$  和  $Q^-(Y|Z) > Q^-(Y)$ ，而在结论中，我们希望有  $\Pr(Y|Z) > \Pr(Y)$ 。这三个正相关的表述其实涉及了三个不同的概率。第一个正相关表述等价于  $\Pr(Y|Z \wedge X) > \Pr(Y|X)$ ，第二个正相关表述等价于  $\Pr(Y|Z \wedge \neg X) > \Pr(Y|\neg X)$ ，因此，他们是完全不同于第三个正相关的。

实际上，并不一定成立！有反直觉的例子：存在  $X, Y, Z$  和对应的随机真值表，使得

- $\Pr(Y|Z \wedge X) > \Pr(Y|X)$ .
- $\Pr(Y|Z \wedge \neg X) > \Pr(Y|\neg X)$ .
- 然而， $\Pr(Y|Z) \leq \Pr(Y)$ .

这样的现象和 *Simpson* 悖论有关。

举个例子，球员甲的两分球和三分球命中率均高于球员乙，但是球员甲的总投篮命中率却可能低于乙。具体来说，有一个班中一半同学来自北京大学，另一半来自清华大学，我们抽出一名同学 Bob，估计 Bob 投篮命中的概率。我们给如下记号：

- 记  $Y$  为 Bob 投篮命中，记  $X$  为 Bob 投出一个两分球，则  $\neg X$  为 Bob 投出一个三分球（我们这里只考虑有两分和三分球）。
- 记  $Z$  为 Bob 来自清华大学。
- $\Pr(Y)$  表示全班学生的投篮命中率。
- $\Pr(Y|Z)$  表示全班来自清华大学的学生的投篮命中率。
- $\Pr(Y|X)$  表示全班学生的两分命中率。
- $\Pr(Y|\neg X)$  表示全班学生的三分命中率。
- $\Pr(Y|Z \wedge X)$  表示这个班来自清华大学的学生的两分命中率。
- $\Pr(Y|Z \wedge \neg X)$  表示这个班来自清华大学的学生的三分命中率。

考虑这样一个投篮数据的实例：

	全班同学	来自清华大学
两分球	50/100	6/10
三分球	1/101	1/100
总命中率	51/201	7/110

$\Pr(Y) = 51/201$ ,  $\Pr(Y|Z) = 7/110$  为投篮命中率.  $\Pr(Y|X) = 50/100 = 1/2$ ,  $\Pr(Y|Z \wedge X) = 6/10 = 3/5$  为两分命中率.  $\Pr(Y|\neg X) = 1/101$ ,  $\Pr(Y|Z \wedge \neg X) = 1/100$  为三分命中率. Simpson 悖论在这一实例下表现为：这个班里来自清华大学的学生两分命中率和三分命中率分别都比全班平均水平高，但总体投篮命中率反倒比全班水平低。

我们将这两个概率用全概公式展开来寻找原因：

$$\Pr(Y|Z) = \Pr(Y|Z \wedge X) \Pr(X|Z) + \Pr(Y|Z \wedge \neg X) \Pr(\neg X|Z),$$

$$\Pr(Y) = \Pr(Y|X) \Pr(X) + \Pr(Y|\neg X) \Pr(\neg X).$$

上下两行对应的项， $\Pr > \Pr$ . 然而，关键是有可能发生  $\Pr(X|Z) \neq \Pr(X)$ . 在上面篮球的例子中表现为清华大学的同学选择投两分球和三分球的比例和全班同学不同。这表明，Simpson 悖论发生的核心原因与基率谬误类似，即没有正确区分先验概率和后验概率。

接下来我们考虑合取谬误（见本章开头）。合取谬误是一种认知偏差。我们这里呈现一个简化的例子作为讨论。Linda 是一位单身、外向且年龄为 31 岁的女性。在大学期间，她主修哲学，十分关注种族歧视和社会公正问题，而且曾参加过反核游行。（记为  $E$ ）请问以下哪一件事情更可能发生？

1. Linda 是一名银行出纳员（记为  $B$ ）。
2. Linda 是一名银行出纳员，同时她还积极参加女权运动（记为  $B \wedge F$ ）。

在调查实验中，多数被试选择了 2。但是，我们可以肯定  $\Pr(B \wedge F|E) \leq \Pr(B|E)$ 。如何理解这一现象？

为了理解这种谬误产生的原因，考虑合取原则：如果  $E$  是  $P \wedge Q$  的“好论据”，那么  $E$  也是  $P$  的“好论据”，也就是说，如果我可以论证两件事同时成立，那么我也一定可以论证其中一件成立。在演绎推理中，因为  $E \rightarrow (P \wedge Q) \vdash E \rightarrow P$ ，所以合取原则成立。

类似确凿性原则，在合情推理中，合取原则也可以从证据支持和正相关性两方面考虑。从证据支持的角度，合取原则成立，这是因为，如果  $\Pr(P \wedge Q|E) > 0.5$ ，那么  $\Pr(P|E) \geq \Pr(P \wedge Q|E) > 0.5$ 。

然而，从正相关性的角度，合取原则未必成立。也就是说，假设  $\Pr(P \wedge Q|E) > \Pr(P \wedge Q)$ ，不一定能推出  $\Pr(P|E) > \Pr(P)$ 。当人们给定对 Linda 的描述  $E$  的时候，很容易建立起  $E$  和  $B \wedge F$  的正相关性，然而这并不意味着  $E$  和  $B$  是正相关的！因此发生了合取谬误。从 Simpson 悖论和合取谬误可以看出，只依靠正相关性进行推理很容易犯错误，因此证据支持（极大似然）是归纳强论证不可缺少的要素。

## §1.4 先验模型的存在性

历史上，绝大部分成功的科学理论都是依赖实验与观察数据的，我们就以家喻户晓的 Newton 万有引力定律为例。这一定律表明，两个物体之间的引力与他们的质量分别成正比，与他们的距离的平方成反比。我们总是听到那个美妙的故事，说 Newton 因为被苹果砸到才发现了这一定律。然而，这个故事忽略了这一定律背后的伟大科学家，Kepler，他总结老师 Tycho 的海量肉眼观测数据，提出了行星运动的三定律。没有这些定律，Newton 也不可能准确写出平方反比定律。

科学的原则总是提出理论，最后再用实验去验证或者证伪这样的理论。几乎所有的理论都源自对实验数据的观察总结，我们将这种科学理论称之为数据驱动的科学理论。几乎同样的技术和想法被广泛应用在深度学习中，最大的区别是，科学家，或者说人类，被复杂的人工神经网络代替。

不论是科学理论还是深度学习，数据驱动方式的核心假设是，确实存在一个世界的先验模型，我们通过实验数据来逼近这个模型。本节将要给出一个重要的定理，*de Finetti* 定理，它表明在一定条件下，无限多的实验数据本身一定蕴含了一个先验模型。因此，数据驱动的模式是合理的，这也成为了 Bayes 概率论的基石之一。

接下来，我们建模这一问题. 首先是数据，实验观测数据需要遵循可重复性原则. 假如我们观察到了数据  $x_1, x_2, \dots, x_n$ ，由于实验有不可控因素，实验结果会有随机性，所以我们假设  $X_1, \dots, X_n$  是观测结果  $x_1, \dots, x_n$  所对应的随机变量. 从合情推理的角度来说，随机性更像是人对于实验不确定性的一种认识或信念. 所谓的可重复就是指，这些实验不管以什么样的顺序完成，他们都应该产生相同的结果. 假设  $\pi$  是一个  $\{1, 2, \dots, n\}$  的排列，我们要求

$$(X_1, \dots, X_n) \stackrel{d}{=} (X_{\pi(1)}, \dots, X_{\pi(n)}).$$

也就是说，考虑了随机性之后，打乱实验的顺序，观测结果的概率分布应该是不变的. 随机变量列这样性质的被称可交换性.

那么，什么是理论模型呢？一个定量的理论模型可以看成某种对真实世界的参数化建模，例如参数为  $\theta$ ，它能够解释  $X_i$  中的某些规律. 如果我们知道了真实世界的  $\theta$ ，那么观测结果  $X_i$  之间不仅仅是可交换的，而且是独立的. 这里的意思是，科学理论所能解释的部分应该使得  $X_i$  的不确定性不再依赖于其他的  $X_j$ ，而仅仅在于一次实验内部. 这样的性质被称为条件独立性.

接下来的问题是，我们是否可以发现一个理论模型？更直接的问题是，如果给我们足够多的数据，我们是否能够直接形成一个对于  $\theta$  的先验信念（即概率密度  $p_\theta$ ），而不需要再人为有某种特别的先验信念？这就是 de Finetti 定理要回答的问题.

现在我们正式陈述这一定理. 首先，让实验次数为无穷，即  $X_1, X_2, \dots$ . 此时，可交换性指的是任意有限个  $X_{i_1}, X_{i_2}, \dots, X_{i_n}$  以及任意一个  $\{i_1, \dots, i_n\}$  上的排列  $\pi$  都有

$$(X_{i_1}, X_{i_2}, \dots, X_{i_n}) \stackrel{d}{=} (X_{\pi(i_1)}, X_{\pi(i_2)}, \dots, X_{\pi(i_n)}).$$

此时，我们有 de Finetti 定理：

**定理 1.6 (de Finetti 定理)** 如果  $X_1, X_2, \dots$  是可交换的连续型随机变量列，那么存在一个概率密度  $p_\theta$ ，使得对于任意的  $n$  和任意的  $x_1, \dots, x_n$ ，都有

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = \int p_\theta(\theta) \prod_{i=1}^n p_{X_i}(x_i|\theta) d\theta.$$

我们来解释这一定理与上面讨论的对应. 如果我们知道了  $\theta$ ，那么  $X_1, \dots, X_n$  条件在  $\theta$  上的条件密度是  $\prod_{i=1}^n p_{X_i}(x_i|\theta)$ ，因而是条件独立的，这符合了我们对于理论模型的基本要求. 另外，如果我们不知道  $\theta$ ，只知道无穷多的实验数据，我们并不需要人为再给定一个关于  $\theta$  的先验，数据本身就已经足够形成先验了，不需要人为的先验.

## §1.5 章末注记

[lhy: TODO]

Euclid 在他伟大的《几何原本》<sup>2</sup>最早尝试将数学整理为一个完整的公理系统.

---

<sup>2</sup>实际上, Euclid 的书名直译是“原理”, 并没有“几何”一词. 这说明 Euclid 其实在探求某些更加本质的东西, 这样的思想后来被发展为了我们今天的形式推理系统或公理系统.

## 第二部分

### 信息与数据

## 第三部分

### 决策与优化



## 第四部分

### 逻辑与博弈

## 第五部分

### 认知逻辑

## 第六部分

### 附录：预备知识

## 参考文献

- [Bre57] Leo Breiman. The Individual Ergodic Theorem of Information Theory. *The Annals of Mathematical Statistics*, 28(3):809–811, 1957.
- [CT12] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 2012.
- [Huf52] David A. Huffman. A Method for the Construction of Minimum-Redundancy Codes. *Proceedings of the IRE*, 40(9):1098–1101, September 1952.
- [Inf] Information | Etymology, origin and meaning of information by etymonline. <https://www.etymonline.com/word/information>.
- [Jay02] Edwin T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2002.
- [KL51] S. Kullback and R. A. Leibler. On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.
- [LLG<sup>+</sup>19] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension, October 2019.
- [McM53] Brockway McMillan. The Basic Theorems of Information Theory. *The Annals of Mathematical Statistics*, 24(2):196–219, June 1953.
- [RHW86] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations*, pages 318–362. MIT Press, Cambridge, MA, USA, January 1986.

- [Rob49] Robert M. Fano. *The Transmission of Information*. March 1949.
- [Sha48] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, July 1948.
- [Shi96] A. N. Shiryaev. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer, New York, NY, 1996.
- [Tin62] Hu Kuo Ting. On the Amount of Information. *Theory of Probability & Its Applications*, 7(4):439–447, January 1962.
- [Uff22] Jos Uffink. Boltzmann’s Work in Statistical Physics. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2022 edition, 2022.
- [李 10] 李贤平. 概率论基础. 高等教育出版社, 2010.

# 索引

Bayes 概率论, 14

de Finetti 定理, 21, 22

Hilbert 推理系统, 6

Simpson 悖论, 19

三段论

弱  $\sim$ , 8, 15

强  $\sim$ , 8, 15

事件域, 12

似然, 9, 14

偏函数, 9

先验概率, 14

公理系统, 23

前提, 5-7

可交换性, 22

可重复性原则, 22

合取原则, 21

合取谬误, 3, 20

合情推理, 3, 8

与规则, 11

否定规则, 11

等值原则, 9

后验概率, 14

命题, 3

命题公式, 3

命题变元, 3

命题逻辑, 3

基率, 14

基率谬误, 14

完备性定理, 7

归纳强论证, 16, 17

归谬法, 6

形式推理系统, 3, 6, 23

推导规则, 6

最大似然, 15

有效论证, 15

条件独立性, 22

概率, 12

正相关, 16

演绎推理, 3, 7

相关性, 17

确凿性原则, 19

无条件  $\sim$ , 19

等值, 7

结论, 5-7

肯定前件, 6

自指, 3

自然演绎系统, 6

认可度, 17

认可度似然比, 17

认可概率增量, 17

证据支持, 16

语义, 4

语义蕴含, 5

语形, 4

赋值, 4

连接词, 3, 4

逻辑结论, 15

重言式, 7

随机真值表, 16