《因果论》第一章汇报：概率、图及因果模型

梁克维　张逸凡

浙江大学数学科学学院

2024 年 4 月 7 日

概率的贝叶斯解释

根据概率描述对事件的信念程度，并使用数据来增强、更新或削弱这些信念程度.

- 选用 $P(A|B)=P(A)$ 作为命题 A 与 B独立的定义.
  (而非 $P(A,B)=P(A)P(B)$)
- 选用 $P(A|B,C)=P(A|C)$ 作为 A 与 B 在给定 C 时条件独立的定义.
  (记作 $(A \perp\!\!\!\perp B|C)_P$)
- 容易导出链式法则公式：

$$P(E_1,E_2,...,E_n)=P(E_n|E_{n-1},...,E_2,E_1)...P(E_2|E_1)P(E_1) \tag{1}$$

- 贝叶斯推断的核心在于**反演公式**：

$$P(H|e)=\frac{P(e|H)P(H)}{P(e)} \tag{2}$$

贝叶斯主义者将 (2) 式看作证据更新信念的标准规则.

- 将 (2) 式除以其互补式 $P(H^c|e)$，可以得到

$$\frac{P(H|e)}{P(H^c|e)}=\frac{P(e|H)}{P(e|H^c)}\frac{P(H)}{P(H^c)} \tag{3}$$

其中等号左侧的分式称为**后验概率**，等号右侧的两个分式分别称为**似然比**和**先验概率**.

概率图模型

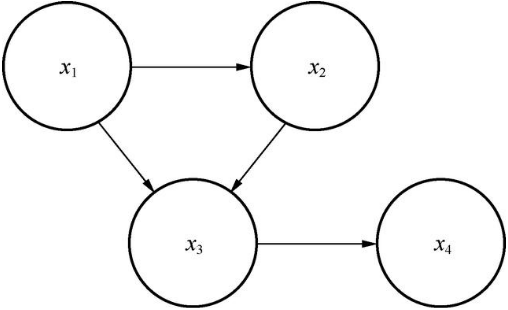

图 1: 一个概率图

- 图由顶点集 V 和边集 E 组成，其中顶点对应于 (随机) 变量，边表示变量对之间的某种关系.
- 边可以是有向的也可以是无向的. 在某些应用中，我们使用双向的虚线表示未观察到的共同原因 (混杂因子).
- 若所有的边都是有向的，我们得到有向图. 不包含有向环的图称为无环图. 既有向又无环的图称为有向无环图 (DAG).

贝叶斯网络

有向图，尤其是 DAG，常用于表示因果关系或时间关系，这种图称为贝叶斯网络.
由 (1) 式，我们有 $P(x_1, \ldots, x_n) = \prod_j P(x_j | x_1, \ldots, x_{j-1})$.

现假设 X_j 只对其中小部分前驱变量敏感，记这部分前驱变量为 PA_j，那么上式可以写为：

$$P(x_1, \ldots, x_n) = \prod_j P(x_j | pa_j) \tag{4}$$

我们称 PA_j 为 X_j 的马尔可夫父代集合.

构造 DAG 的递归方法

- 首先，假定我们由经验得知 $X_1, ..., X_n$ 服从一定的时间或因果顺序.
- 从节点对 (X_1, X_2) 开始，当且仅当这两个变量相关时，画一个从 X_1 到 X_2 的箭头.
- 继续处理 X_3，若 X_3 独立于 X_1, X_2，则不画箭头. 否则，检测 X_2 是否使 X_3 条件独立于 X_1. 若该条件成立，则画一个从 X_2 到 X_3 的箭头. 对于 X_1 使 X_3 条件独立于 X_2 的情况类似处理. 若无条件独立，则从 X_1 和 X_2 各画一个箭头指向 X_3.
- 一般地，在构造的第 j 阶段，选择 X_j 的马尔可夫父代集合 PA_j 并从 PA_j 的每一个成员画一个箭头指向 X_j.

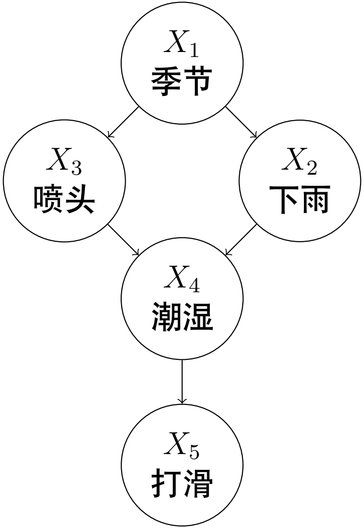

图 2: 一个由递归方法构造出的贝叶斯网络

马尔可夫相容性及其充要条件

若概率函数 P 容许有向无环图 G 有形如 $P(x_1,...,x_n)=\prod_j P(x_j|pa_j)$ 的分解，那么我们认为 P 与 G马尔可夫相容.

- (有序马尔科夫条件) 概率分布 P 与有向无环图 G 马尔可夫相容的充分必要条件是：以 G 中每个变量的父代变量为条件时，该变量与其所有前驱节点独立.
- (马尔可夫父代条件) 概率分布 P 与有向无环图 G 马尔可夫相容的充分必要条件是：以每个节点的父代节点为条件时，该节点独立于 G 中除了它本身以外所有的非后代节点.

这两个马尔可夫相容性的充要条件的证明在书后的参考文献中给出.

概率图中的三种结构

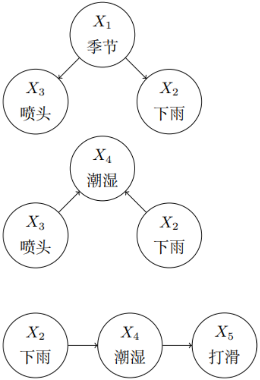

图 3: 概率图三种结构

尾对尾结构

头对头结构

头对尾结构

d-分离准则

考虑一个有向无环图 G 中 3 个不相交的变量集 X, Y, Z. 变量集 Z 将 X 和 Yd-分离(记作 $(X \perp\!\!\!\perp Y|Z)_G$)，是指任何与 G 马尔可夫相容的分布 P 中，$X$ 在 Z 的条件下与 Y 独立.

任意一个从 X 到 Y 的路径 p 被 Z 阻断当且仅当以下两个命题之一成立：

- p 包含了一个头对尾结构或尾对尾结构，而其中间节点在 Z 中.
- p 包含一个头对头结构，而其中间节点不在 Z 中.

变量集 X 与变量集 Y 被变量集 Zd-分离当且仅当所有从变量集 X 到变量集 Y 的路径被变量集 Z 阻断.

DAG 的等价

两个 DAG**观察等价**，是指与其中一个 DAG 马尔可夫相容的概率分布也一定与另外一个 DAG 马尔可夫相容.

- 两个 DAG 观察等价，当且仅当它们拥有相同的骨架及相同的头对头结构.

这说明，对于概率分布 P 而言，与其马尔可夫相容的 DAG 并非唯一的. 若不借助操纵性实验或时间信息，就无法区别两个观察上等价的网络.

干预预言

联合分布告诉我们事件发生的可能性有多大，以及相应的概率如何随着后续的观察而变化，但无法告诉我们相应的概率如何随着后续的干预而变化.

注意区别这两个概念：
若我们观察到喷头打开，那么很可能说明空气湿度本身很低；但若我们干预了喷头使其打开，则不能说明这一点.

然而，因果模型可以告诉我们相应的概率如何随着后续的干预而变化.

若我们实施干预“打开喷头”，则可以得到如下网络：

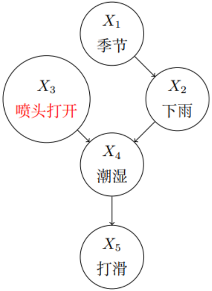

图 4: 干预“打开喷头”的网络

它具有联合分布

$$P_{X3=open}(x_1, x_2, x_4, x_5) = P(x_1)P(x_2|x_1)P(x_4|x_2, X_3 = open)P(x_5|x_4) \tag{5}$$

因果贝叶斯网络

令 $P(v)$ 表示变量集 V 上的概率分布，$P_x(v)$ 表示设置变量子集 X 为常数 x 的干预 $do(X=x)$ 后的分布，P_* 表示所有干预分布 $P_x(v)$ 的集合. 有向无环图 G 是与 P_* 相容的因果贝叶斯网络，当且仅当如下三个条件对任意 $P_x(v) \in P_*$ 都成立：

- $P_x(v)$ 与 G 马尔可夫相容；
- 对于所有的 $V_i \in X$，当 v_i 与 $X=x$ 一致时，$P_x(v_i)=1$；
- 对于所有的 $V_i \notin X$，当 pa_i 与 $X=x$ 一致时，$P_x(v_i|pa_i)=P(v_i|pa_i)$.

这样的定义容许我们计算出干预 $do(X=x)$ 后的结果分布：
对于所有与 x 一致的 v，

$$P_x(v) = \prod_{\{i|V_i \notin X\}} P(v_i|pa_i) \tag{6}$$

于是，(5) 式给出的联合分布是合理的.

因果关系的稳定性

因果关系比概率关系更“稳定”，本质上是由于因果关系属于“本体论”，而概率关系属于“认识论”.

考虑如下两个命题：

- S_1：打开喷头不会导致下雨
- S_2：喷头的状态独立于下雨的状态

应用 d-分离准则，我们知道当我们获知季节的取值时，S_2 由假变为真；而当我们又获知路面是否湿滑时，S_2 又由真变为假. 然而，S_1 始终为真，不论我们获知多少信息.

结构方程

一个函数因果模型由以下形式的一组方程构成：

$$x_i = f_i(pa_i, u_i), i = 1, ..., n \tag{7}$$

其中，U_i 表示由遗漏因子所产生的误差或扰动.

如 (7) 式的一组方程称为结构模型. 若每个变量都有唯一的一个方程，且这个变量出现在方程的左边，那么这个模型成为结构因果模型，或因果模型.

需要注意的是，结构方程中，变量不可以随意移项，这与代数方程有所区别. 这是因为，等式左侧表示因变量，等式右侧表示自变量. 若移项，则结构方程所表示的因果关系颠倒.

因果模型中的概率预测

已知一个如式 (7) 的因果模型. 如果从 PA_j 的每个成员画一个指向 X_i 的箭头，得到的图 G 称为因果图. 若该图是无环的，且误差项是联合独立的，则称该模型为马尔可夫模型.

- (因果马尔可夫条件) 每个马尔可夫模型都可以推导出一个分布 $P(x_1, ..., x_n)$，该分布与 M 对应的因果图 G 马尔可夫相容.

函数模型中的干预

当一个干预将固定值 x 实施到变量子集 X 上时，这些变量子集所对应的方程改写为 $X = x$. 并且，对于所有马尔可夫父代与 X 有交集的变量所对应的方程，X 的取值固定为 x.

修正后的模型所推导出的联合概率函数与 (6) 式相同.

函数模型中的反事实

所谓反事实，就是一种假想的事实，与实际发生的事件相反.
例如，一个人接受了治疗并死亡. 那么我们自然会关心，他的死亡是由治疗导致的，还是与治疗无关？即如果他并没有接受治疗，仍然会死亡吗？

若用普通的统计检验方法去回答反事实的问题是不可能的，因为反事实从未真正发生过. 因此，很多统计学家将反事实的问题视为玄奥的问题.
但是，在日常生活中充满了反事实的表达，因此很多时候我们很有必要去回答反事实的问题. 实际上，反事实问题会在后续章节 (第七章) 中给出更加详尽的解决方法. 在此通过一个例子解释为什么贝叶斯网络中描述的信息不足以回答反事实的问题.

考虑一对独立的随机变量 X 和 Y.X=1 代表治疗，X=0 代表不治疗；Y=1 代表康复，Y=0 代表死亡. 考虑如下两个函数模型：

- 令 $x = u_1, y = u_2$，其中 U_1 与 U_2 独立，服从 $b(1, \frac{1}{2})$
- 令 $x = u_1, y = xu_2 + (1-x)(1-u_2)$

容易验证，在这两个模型中，X 均与 Y 独立，且对于所有的 x 和 y，$P(x, y) = 0.25$. 它们对应同一个因果贝叶斯网络：孤立的 X 点与 Y 点.

然而，模型 1 的含义是，抛一枚硬币，一面对应康复，一面对应死亡，与治疗无关，因此原本死亡的人接受了治疗仍然死亡，原本康复的人接受了治疗仍然康复，反事实与事实相同.

与之相反，模型二的意义是，抛一枚硬币，代表本身康复还是死亡；而治疗的疗效很奇特：本应康复的人治疗后会死亡，本应死亡的人治疗后会康复. 因此反事实与事实截然相反.

由此可见，在这个例子中，我们无法从因果贝叶斯网络提供给我们的信息中回答反事实的问题. 然而，当我们获知结构方程的时候，回答这个问题就很简单了.