目录

文件名

第一部分　条件概率

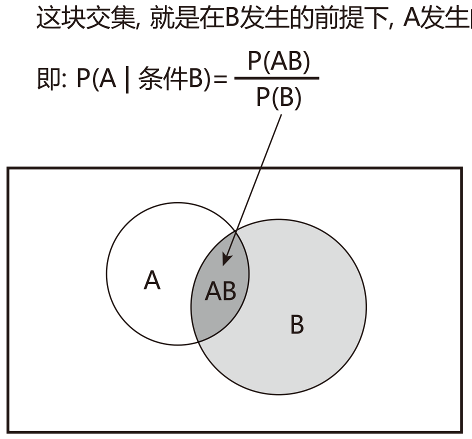
1　“条件概率”：

$$\boxed{P(A|\text{条件}B) = \frac{P(A\cap B)}{P(\text{条件}B)} = \frac{\text{事件}A,B\text{同时发生的概率}}{\text{事件}B\text{发生的概率}}}$$

条件概率是：有A, B 两个事件，和样本空间Ω．其中 $P(B) > 0$, 则，在B已经发生的条件下，A发生的概率，就叫做A对B 的“条件概率”．记作：P(A— 条件B)，读作“在B发生的条件下，A发生的概率”．

即，条件概率公式是：$P(A\ |conditionB) = \frac{\overbrace{\text{在B发生条件下,A发生的样本点数}}^{\text{这个分子即: AB同时发生了}}}{\text{B里面有多少个样本点}} = \frac{\mathrm{n_{AB}}}{\mathrm{n_B}}$

还可写成：$P(A\ |conditionB) = \frac{\mathrm{P\,(A\cap B)}}{\mathrm{P\,(B)}} = \frac{\frac{\mathrm{n_{AB}}}{\mathrm{n}}}{\frac{\mathrm{n_B}}{\mathrm{n}}} = \frac{\mathrm{n_{AB}}}{\mathrm{n_B}}$



如上图所示，注意：概率是个比值，所以你光有分子那块的交集值，是没用的，它还需要与另一个数(分母)去比．

上面公式中，P(AB) (← 也可写作 $P(A\cap B)$) 的计算公式是什么呢？
- 如果事件A，和事件B 是相互独立的，则 $P(AB) = P(A)\cdot P(B)$
- 如果事件A，和事件B 不相互独立，则只能用“条件概率”公式，来求P(AB)，即：

$$\boxed{P(AB) = P(B)\cdot P(A|B) = P(A)\cdot P(B|A)}$$

注意：“条件概率”，和“分步骤法”的区别：
- 分步骤法 (用乘法)：前后每一步骤的事件是相互独立的，彼此没有条件关系．
比如，第一步你结婚，第二步我结婚．我们这两件事发生的概率互不影响．

- 条件概率 (里面也有用到乘法): 前面的事件, 有可能会(但并不一定)影响到后面事件的发生概率. 即前后事件之间并不互相独立.
会影响的例子: 比如一共有100个上岸机会, 则第一步你上岸的成功概率, 会影响到第二步我上岸的成功概率. (你若成功, 留给我的名额数量就会更少.)
彼此独立的例子: 比如在你回国的条件下, 我出门的概率. 两者发生的概率毫无关系. 你回不回国, 跟我会出不出门没半毛钱关系.

例

有6个球, 各有编号. 我们先定义下这些事件:
- B: 取到偶数编号的球
- A_1: 取到1号球
- A_2: 取到2号球
- A_5: 取到大于4号的球

则:
- $\overbrace{\mathrm{P}(\mathrm{A}_1)}^{取到1号球的概率}=\frac{\overbrace{\mathrm{C}_1^1}^{1号球选1}}{\underbrace{\mathrm{C}_6^1}_{全6选1}}=\frac{1}{6}=0.166667$

- $\mathrm{P}(\mathrm{A}_1|\mathrm{B})=\frac{在B条件里面,取到A_1(即1号球)}{B:\ 取到偶数编号的球}=\frac{\overbrace{0}^{偶数编号的球里面,取不到奇数编号的球}}{\underbrace{\mathrm{C}_3^1}_{3个偶数球里面取1个}}=0$

- $\mathrm{P}(\mathrm{A}_2|\mathrm{B})=\frac{\overbrace{\mathrm{C}_1^1}^{1个编号2的球里面,取1个}}{\underbrace{\mathrm{C}_6^3}_{3个偶数球里面取1个}}=\frac{1}{3}$

- $\mathrm{P}(\mathrm{A}_5|\mathrm{B})=\frac{在B条件里面,取到大于4号的球}{B:\ 取到偶数编号的球}=\frac{\overbrace{1}^{5,6号与偶数的交集,只有6号一个球}}{3}$

例

某地10万人, 其中平均有40人患甲病; 有34人患乙病; 两种病都患的有32人. 则, 我们设:
→ 事件A: 是一个人得甲病
→ 事件B: 是一个人得乙病

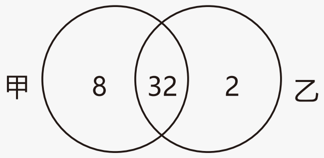
即有:
$P(A)=\frac{40}{100000}=0.0004$
$P(B)=\frac{34}{100000}=0.00034$
$P(AB)=\frac{32}{100000}=0.00032$

根据公式: $P(AB)=P(A)\cdot\underbrace{P(B|A)}_{=\frac{P(AB)}{P(A)}}=P(A)\cdot\frac{P(AB)}{P(A)}$,

$$\leftarrow 即\ P(B|A)=\frac{P(AB)}{P(A)}=\frac{\frac{32}{100000}}{\frac{40}{100000}}=0.8$$

根据公式: $P(AB)=P(B)\cdot P(A|B)=P(B)\cdot\frac{P(AB)}{P(B)}$

$$\leftarrow 即\ P(A|B)=\frac{P(AB)}{P(B)}=\frac{\frac{32}{100000}}{\frac{34}{100000}}=0.941176$$

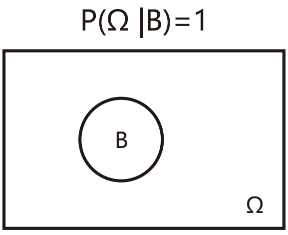


---

2 条件概率的性质

2.1 性质：$P(A|条件B) >= 0$

2.2 性质：$P(\Omega|条件B) = 1$



2.3 性质：$\mathbf{P}\left(\mathbf{A}_1 \cup \mathbf{A}_2\ |\mathbf{B}\right) = \mathbf{P}\left(\mathbf{A}_1\ |\mathbf{B}\right) + \mathbf{P}\left(\mathbf{A}_2\ |\mathbf{B}\right) - \mathbf{P}\left(\mathbf{A}_1\mathbf{A}_2\ |\mathbf{B}\right)$

2.4 性质：$\mathbf{P}\left(\mathbf{A}\ —\mathbf{B}\right) = 1 - \mathbf{P}\left(\overline{\mathbf{A}}\ |\mathbf{B}\right)$

2.5 性质：可列可加性：若$A_1, A_2, ...A_n, ...$ 是“互不相容”的事件，则有：$P(\sum_{i=1}^{\infty} A_i|B) = \sum_{i=1}^{\infty} P(A_i|B) \leftarrow$ 即：“和的概率”，等于“概率的和”

---

3 “条件概率”的乘法公式：

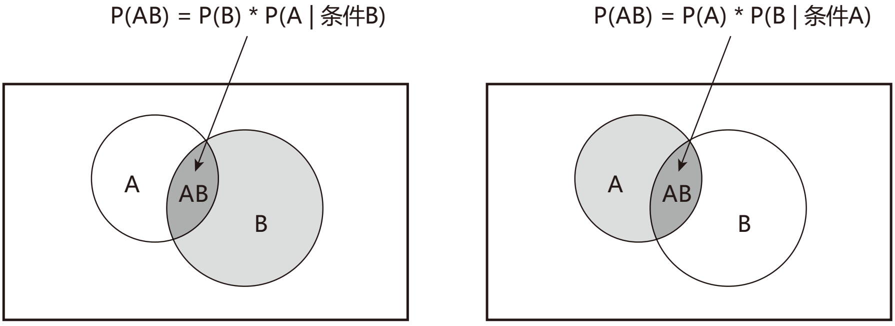
$$\boxed{P(前后) = P(后) \cdot P(前|后) = P(前) \cdot P(后|前)}$$

推导过程:

$$\text{因为}\begin{cases} \mathrm{P(A|B)} = \dfrac{\mathrm{P(AB)}}{\mathrm{P(B)}} \to \text{即 } \mathrm{P(AB)} = \underbrace{\mathrm{P(B)} \cdot \mathrm{P(A|B)}}_{\text{乘法，是交集}\cap\text{的概念}} \quad ① \\ \mathrm{P(B|A)} = \dfrac{\mathrm{P(AB)}}{\mathrm{P(A)}} \to \text{即 } \mathrm{P(AB)} = \underbrace{\mathrm{P(A)} \cdot \mathrm{P(B|A)}}_{\text{乘法，是交集}\cap\text{的概念}} \quad ② \end{cases}$$

①和②，就是“乘法公式”. 即 $\mathrm{P(AB)} = \mathrm{P(B)} \cdot \mathrm{P(A|B)} = \mathrm{P(A)} \cdot \mathrm{P(B|A)}$

其中 $\mathrm{P(A)} > 0$，$\mathrm{P(B)} > 0$



同理, 多个事件的乘法公式就是:

$\rightarrow \boxed{P(ABC) = \underbrace{P(A)} \cdot \underbrace{P(B—A)} \cdot \underbrace{P(C—BA)}}$

↑ 上面“从右往左”看, 就是按 A,B,C 的顺序

$$\rightarrow P(A_1A_2\cdots A_n) = P(A_1) \cdot P(A_2A_1) \cdot P(A_3A_2A_1) \cdot ... \cdot P(\overbrace{A_n}^{\text{即最后一个数字n写在前面.}} \mid \overbrace{A_{n-1}}^{\text{后面的条件, 最后一个数字是: n少一位}} ...A_2A_1)$$

↑ 上面“从右往左”看, 就是按$A_1, A_2, ..., A_n$的顺序

例

有100件产品, 次品率=10%, 即有10件次品. 做不放回抽样, 问: 第3次才取到合格品的概率是?

我们先令:

- A_1 表示第1次取, 就取到了合格品
- A_2 表示第2次取, 取到了合格品
- A_3 表示第3次取, 取到了合格品

那么第3次才取到合格品, 就是:

$$\underbrace{P\left(\overline{A_1}\,\overline{A_2}A_3\right)}_{\text{分步骤来做}} = \underbrace{P\left(\overline{A_1}\right)}_{\text{第1次取,取到次品}} \cdot \underbrace{P\left(\overline{A_2}\,|\overline{A_1}\right)}_{\text{在第1次取到次品的条件下, 第2次取到也是次品}} \cdot \underbrace{P\left(A_3\,|\overline{A_2}\,\overline{A_1}\right)}_{\text{在第1,2次取到次品的条件下, 第3次取到正品}}$$

$$= \frac{\overbrace{C_{10}^1}^{\text{第一步, 10次品里面取1}}}{C_{100}^1} \cdot \frac{\overbrace{C_9^1}^{\text{第二步, 因为第一步已经拿走1件次品, 这第二步就只能在9件次品里面取1件次品了}}}{C_{99}^1} \cdot \frac{\overbrace{C_{90}^1}^{\text{既然第三步要拿到正品, 我们就从90件正品里面取1了}}}{C_{98}^1}$$

$$= \frac{10}{100} \cdot \frac{9}{99} \cdot \frac{90}{98} = 0.00834879$$

例

某产品:

- 甲公司占60%市场份额, 且其产品合格率是 90%
- 乙公司占40%市场份额, 且其产品合格率是 80%

我们先定义下这些事件:

- J: 表示产品是甲的
- $\overline{J}$: 表示产品是乙的
- Q (qualified) : 表示产品是“合格”的

- $\overline{\mathrm{Q}}$: 表示产品是“不合格”的

问, 你买一个产品, 是甲公司的, 并且是合格的概率是?
$\mathrm{P}(\mathrm{JQ})=\underbrace{\mathrm{P}(\mathrm{J})}_{=0.6}\cdot\underbrace{\mathrm{P}(Q\mid J)}_{甲的合格率=0.9}=0.54$

问, 你买一个产品, 是乙公司的, 并且是合格的概率是?
$\mathrm{P}(\bar{\mathrm{J}}\mathrm{Q})=\underbrace{\mathrm{P}(\bar{\mathrm{J}})}_{=0.4}\cdot\underbrace{\mathrm{P}(Q\mid\bar{\mathrm{J}})}_{乙条件下的合格率=0.8}=0.32$

例

抽签, 共10签, 其中有4个为“成功上岸”的好签. 甲乙丙三人, 按顺序依次去抽, 不放回.
我们先设定事件:
- A: 表示甲抽到“成功”
- B: 表示乙抽到“成功”
- C: 表示丙抽到“成功”

问, (1) 甲抽到“成功”的概率? $\mathrm{P}(\mathrm{A})=\frac{\mathrm{C}_{4好签}^{1}}{\mathrm{C}_{10签}^{1}}=\frac{4}{10}=0.4$

(2) 甲乙都抽到“成功”的概率?

$$\mathrm{P}(\mathrm{AB})=\underbrace{\mathrm{P}(\mathrm{A})}_{第1步:甲先成功}\cdot\underbrace{P(B|A)}_{第2步:在甲成功的前提下,\ 乙再成功}$$
$$=\frac{\overbrace{\mathrm{C}_{4好签}^{1}}^{甲先抽掉一张好签}}{\mathrm{C}_{10签}^{1}}\cdot\frac{\overbrace{\mathrm{C}_{4好签-1}^{1}}^{乙就只能从剩下的3张好签中来抽了}}{\mathrm{C}_{10签-1}^{1}}=\frac{4}{10}\cdot\frac{3}{9}=0.133333$$

(3) 甲失败, 乙成功的概率?

$$P(\overline{A}B)=\underbrace{P(\overline{A})}_{第1步:甲先失败}\cdot\underbrace{P(B\mid\overline{A})}_{第2步:在甲失败的前提下,\ 乙再成功}$$
$$=\frac{\overbrace{\mathrm{C}_{6坏签}^{1}}^{甲先从共6张坏签中取1}}{\mathrm{C}_{10签}^{1}}\cdot\frac{\overbrace{\mathrm{C}_{4好签}^{1}}^{乙从共4张好签中取1}}{\mathrm{C}_{10签-1}^{1}}=\frac{6}{10}\cdot\frac{4}{9}=0.266667$$

(4) 甲乙丙都抽到“成功”的概率?

$$P(ABC)=\underbrace{P(A)}_{第1步:甲先成功}\cdot\underbrace{P(B\mid A)}_{第2步:在甲成功的前提下,\ 乙再成功}\cdot\underbrace{P(C|BA)}_{第3步:在甲乙都成功的前提下,丙再成功}$$
$$=\frac{\mathrm{C}_{4好签}^{1}}{\mathrm{C}_{10签}^{1}}\cdot\frac{\mathrm{C}_{还剩3好签}^{1}}{\mathrm{C}_{还剩9签}^{1}}\cdot\frac{\mathrm{C}_{还剩2好签}^{1}}{\mathrm{C}_{还剩8签}^{1}}$$
$$=\frac{4}{10}\cdot\frac{3}{9}\cdot\frac{2}{8}=0.0333333$$

例

你的新邻居, 是个 is very shy and withdrawn, invariably helpful but with very little interest in people or in the world of reality. A meek and tidy soul, he has a need for order and structure, ant a passion for detail.
你觉得他更可能是 – 图书管理员, 还是农民?

大多数人会认为他是图书馆管理员. 其实, 这种判断是”非理性”的. 问题的关键在于 –你没有把这两种职业的”人数比例”考虑进去.

→ 在美国,农民与图书馆管理员,两者的数量之比是$20:1$.
根据这个比例,我们可以假设他们人数的概率分别为:

$$\begin{cases} P(\text{农民职业}) = \dfrac{20}{100} \\ P(\text{图书管理员职业}) = \dfrac{1}{100} \end{cases}$$

→ 假如你听到”彬彬有礼”这类描述,你的直觉是:
40%的图书馆管理员符合这个描述,而只有10%的农民符合这个描述.
如果这是你的估计,那就意味着:

$$\begin{cases} \text{在}100\%\text{的纯粹农民群体中,彬彬有礼者的占比概率是:} \\ \qquad P(\text{彬彬有礼 }|\text{农民职业}) = 0.1 \\ \text{在}100\%\text{的纯粹图书管理员群体中,彬彬有礼者的占比概率是:} \\ \qquad P(\text{彬彬有礼 }|\text{图书管理员职业}) = 0.4 \end{cases}$$

→ 在既有农民,又有图书管理员的混合群体中:

$$\begin{cases} \text{在属于农民职业的前提下,是彬彬有礼者的概率是:} \\ P(\text{农民职业} \cap \text{彬彬有礼}) = \underbrace{P(\text{农民职业})}_{\text{第一步,是农民的概率}} \cdot \underbrace{P(\text{彬彬有礼 }|\text{农民职业})}_{\text{第二步,在是农民前提下,彬彬有礼者的概率}} \\ \qquad = 0.2\cdot 0.1 = 0.02 \\ \text{在属于图书管理员职业的前提下,是彬彬有礼者的概率是:} \\ P(\text{图书管理员职业} \cap \text{彬彬有礼}) = P(\text{图书管理员职业}) \cdot P(\text{彬彬有礼 }|\text{图书管理员职业}) \\ \qquad = 0.01\cdot 0.4 = 0.004 \end{cases}$$

→ P(在彬彬有礼者中,取1人为图书管理员的概率)

$$= \frac{P(\text{图书管理员职业}\cap\text{彬彬有礼})}{P(\text{农民职业}\cap\text{彬彬有礼}) + P(\text{图书管理员职业}\cap\text{彬彬有礼})}$$

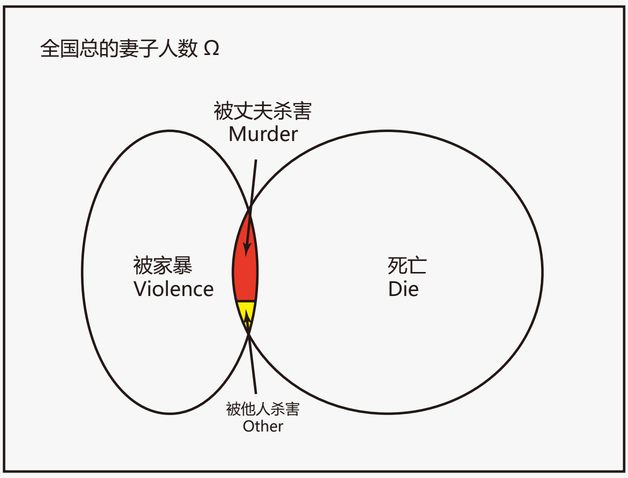
$$= \frac{0.004}{0.02 + 0.004} = 0.166667$$

所以, 即使你认为“符合这个描述的人是一个图书馆管理员的可能性, 是一个农民的4倍”, 也抵不过农民的数量很多.

例

辛普森杀妻案, 原告证明辛普森常常家暴前妻. 他们认为, 长期家暴说明辛普森有杀妻的动机. 被告律师则举出数据反驳说, 美国有400万被家暴的妻子, 但只有1432名被丈夫杀害, 这个概率只有 $\frac{1432}{400\text{万}}$ = 比1/2500还低. 所以家暴证明不了辛普森谋杀.
被告想表达的是: 在“家暴”这个事件前提条件下, 丈夫谋杀妻子的概率不高. 即 P(丈夫家暴 ∩ 丈夫杀妻) = 概率值很低.

你怎么看? 事实上, 被告举出的概率, 不适用于这个案子上. 因为本案的妻子已经死亡, “妻子已死”也变成了一个已经存在的前提条件. 所以现在我们要看的概率就是: P(丈夫家暴 ∩ 妻子已死亡 ∩ 是丈夫杀妻)=? 即: 在“被家暴”且“死亡”的妻子数量里面 (这里就有两个前提条件了, 而不是仅一个前提条件), 有多少是被丈夫杀害的?



即:

- 辛普森律师一方的概率公式是: $\frac{红色}{Violence} < \frac{1}{2500}$
- 妻子一方律师的概率公式是: $\frac{红色}{红色+黄色} = 93\%$

根据美国1992年发布的数据推算: 每10万个被家暴的妇女中, 有43个会被谋杀. 其中40个是被丈夫谋杀, 其他3个是被丈夫以外的人谋杀. 那么, 条件概率就是:

$P(A|B) = \frac{P(AB)}{P(B)}$

$$P(丈夫杀 —家暴 \cap 妻死) = \frac{P(家暴 \cap 妻死 \cap 丈夫杀)}{P(家暴 \cap 妻死)} = \frac{\frac{40}{100000}}{\frac{43}{100000}} = 0.930233$$

你仔细体会一下两者的不同:

- 辛普森方, 是说: 在所有“之后活着和死去”的被家暴的妻子里, 被丈夫杀了的可能性是多大. 即 $\frac{丈夫杀害}{条件: 1.被家暴}$
- 妻子方, 是说: 在所有“死去”的被家暴的妻子里, 被丈夫杀了的可能性是多大? 即$\frac{丈夫杀害}{条件: 1.被家暴\ \&\ 2.死亡}$

不过, 即使概率高达93%，也不能绝对证明辛普森杀了妻子. 因为**“条件概率”只表示统计意义上的“相关性”, 并不代表“因果关系”.** 即只说明: 家暴和谋杀妻子之间有很强的相关性。