例 2.7 随意抛 n 次硬币, 证明正面向上的次数是偶数 (或奇数) 的概率为 $1/2$.

证明 用事件 A 表示前 $n-1$ 次抛硬币正面向上的次数为偶数, 其对立事件 $\bar{A}$ 表示前 $n-1$ 次抛硬币向上的次数为奇数, 事件 B 表示前 n 次硬币向上的次数为偶数. 于是有

$$P(B)=P(A)P(B|A)+P(\bar{A})P(B|\bar{A})=\frac{P(A)}{2}+\frac{P(\bar{A})}{2}=\frac{1}{2}.$$

方法二: 直接计算概率. 若正面向上的次数是偶数, 则随意抛 n 次硬币中正面向上的次数为偶数分别有 $\{0,2,4,\ldots,2k\}$ $(2k\leqslant n)$, 根据概率公式直接计算有

$$\sum_{0\leqslant k\leqslant n/2}\binom{n}{2k}\left(\frac{1}{2}\right)^{2k}\left(\frac{1}{2}\right)^{n-2k}=\frac{1}{2^n}\sum_{0\leqslant k\leqslant n/2}\binom{n}{2k}=\frac{1}{2},$$

这里使用公式 $\sum_{0\leqslant k\leqslant n/2}\binom{n}{2k}=2^{n-1}$.

例 2.8 设有 n 个盒子, 每个盒子有 a 个白球和 b 个红球, 现从第一个盒子取出一球放入第二个盒子, 第二个盒子取出一球放入第三个盒子, 依次类推, 求从最后一个盒子取出一球是红球的概率.

解 用 A_i 表示从第 i 个盒子取出红球的事件 $(i\in[n])$, 则 $\bar{A}_i$ 表示取出白球的事件. 有

$$P(A_1)=b/(a+b)\qquad 和 \qquad P(\bar{A}_1)=a/(a+b)\ .$$

根据全概率公式有

$$P(A_2)=P(A_1)P(A_2|A_1)+P(\bar{A}_1)P(A_2|\bar{A}_1)=\frac{b}{a+b}\times\frac{b+1}{a+b+1}+\frac{a}{a+b}\times\frac{b}{a+b+1}=\frac{b}{a+b}.$$

由此可知 $P(\bar{A}_2)=a/(a+b)$. 依次类推重复上述过程 $n-1$ 次, 最后一个盒子中取出一球是红球的概率为 $b/(a+b)$.

2.2.2 贝叶斯公式

贝叶斯公式是概率论中另一个重要的公式, 探讨在结果发生的情况下由何种原因导致结果. 具体而言, 假设有 $A_1,A_2,\cdots,A_n$ 种原因导致事件 B 发生, 贝叶斯公式研究在事件 B 发生情况下由原因 A_i 导致的概率, 即条件概率 $P(A_i|B)$.

定理 2.3 设 $A_1,A_2,\cdots,A_n$ 是样本空间 Ω 的一个分割, 用 B 表示任一事件且满足 $P(B)>0$. 对任意 $1\leqslant i\leqslant n$ 有

$$P(A_i|B)=\frac{P(A_iB)}{P(B)}=\frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n}P(A_j)P(B|A_j)},$$

该公式被称为 **贝叶斯公式** (Bayes' formula).

贝叶斯公式由条件概率和全概率公式直接推导可得. 由于任何事件和其对立事件都是样本空间的一个分割, 对任意概率非零的事件 A 和 B 有

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

全概率公式和贝叶斯公式应用的条件是相同的, 但解决的问题不同: 将事件 $A_1,A_2,\cdots,A_n$ 看作事件 B 发生的"原因", 而事件 B 是伴随着原因 $A_1,A_2,\cdots,A_n$ 而发生的"结果". 若知道各种原因 $P(A_i)$, 以及在该原因下事件 B 发生的概率 $P(B|A_i)$, 此时利用全概率公式计算结果事件 B 发生的概率; 若结果事件 B 已经发生, 利用贝叶斯公式探讨由原因 A_i 导致该结果的概率 $P(A_i|B)$.

贝叶斯公式被应用于生活中的很多决策问题, 与决策理论密切相关, 下面来看一个简单的例子:

例 2.9 小明在一次 AI 竞赛初期不利. 方法不好的概率为 50%, 改进方法后成功的概率为 50%; 程序有误的概率为 30%, 纠正程序后成功的的概率为 60%; 数据不充分的概率为 20%, 采集更多数据后成功的的概率为 80%. 因时间限制只能选择一种改进方案, 小明应该选择哪一种.

解 用 B 表示小明成功的事件, 用 A_1,A_2,A_3 分别表示方法不好、程序有误、数据不充分三个事件, 根据题意有

$$P(A_1)=50\%,\ P(A_2)=30\%,\ P(A_3)=20\%,\ P(B|A_1)=50\%,\ P(B|A_2)=60\%,\ P(B|A_3)=80\%.$$

根据全概率公式, 小明成功的概率为

$$P(B)=P(A_1)P(B|A_1)+P(A_2)P(B|A_2)+P(A_3)P(B|A_3)=59\%.$$

根据贝叶斯公式有

$$\begin{aligned}
P(A_1|B)&=P(A_1)P(B|A_1)/P(B)=25/59,\\
P(A_2|B)&=P(A_2)P(B|A_2)/P(B)=18/59,\\
P(A_3|B)&=P(A_3)P(B|A_3)/P(B)=16/59,
\end{aligned}$$

因此小明应该选择设计新方法来获得理想排名的概率更高.

例 2.10 (三囚徒问题) 三犯人 a, b, c 均被判为死刑, 法官随机赦免其中一人, 看守知道谁被赦免但不会说. 犯人 a 问看守: b 和 c 谁会被执行死刑? 看守的策略: i) 若赦免 b, 则说 c; ii) 若赦免 c, 则说 b; iii) 若赦免 a, 则以 1/2 的概率说 b 或 c. 看守回答 a: 犯人 b 会被执行死刑. 犯人 a 兴奋不已, 因为自己生存的概率为 1/2. 犯人 a 将此事告诉犯人 c, c 同样高兴, 因为他觉得自己的生存几率为 2/3. 那么谁才是正确的呢?

解 用事件 A, B, C 分别表示犯人 a, b, c 被赦免, 由题意可知

$$P(A)=P(B)=P(C)=1/3.$$

用事件 D 表示看守人说犯人 b 被执行死刑, 则有

$$P(D|A)=1/2,\quad P(D|B)=0\quad \text{和}\quad P(D|C)=1.$$

由全概率公式有

$$P(D)=P(A)P(D|A)+P(B)P(D|B)+P(C)P(D|C)=1/2.$$

由贝叶斯公式有

$$P(A|D)=P(A)P(D|A)/P(D)=1/3\quad \text{和}\quad P(C|D)=P(C)P(D|C)/P(D)=2/3,$$

所以犯人 a 的推断不正确, 犯人 c 的推断正确.

贝叶斯公式给出了一种重要的推理逻辑: 假设 $A_1,A_2,\cdots,A_n$ 是导致事件 B 的"原因", 概率 $P(A_i)$ 被称为 **先验概率** (prior probability), 根据先前的经验(数据)得到. 若现在产生了新证据 B, 条件概率 $\Pr(A_i|B)$ 被称 **后验概率** (posterior probability), 反映了在新证据下对"原因"的新知识. 条件概率 $P(B|A_i)$ 称为 **似然度** (likelihood). 根据贝叶斯公式有

$$\text{后验概率}\, P(A_i|B)=\frac{\text{先验概率}\, P(A_i)}{\text{证据概率}\, P(B)}\times \text{似然度}\, P(B|A_i).$$

贝叶斯公式在人工智能中具有重要的应用, 这里仅仅以医疗辅助诊断为例: 为了诊断病人患有 $A_1,A_2,\ldots$ 中哪一种疾病, 先检查某指标 B (如血糖、血脂、血钙等); 根据以往的数据或经验给出先验概率 $P(A_i)$, 通过医学知识或数据确定概率 $P(B|A_i)$, 通过贝叶斯公式计算后验概率 $P(A_i|B)$. 在实际应用中, 通常会检查多个指标综合所有的后验概率进行诊断.

2.3 事件独立性

条件概率 $P(B|A)$ 考虑在事件 A 发生的情况下事件 B 发生的概率, 通常有 $P(B|A)\neq P(B)$, 即事件 A 发生通常会改变事件 B 发生的可能性. 然而在有些情况下, 事件 A 的发生对事件 B 的发生可能没有任何影响, 即事件的独立性.

2.3.1 两事件的独立性

定义 2.3 设 (Ω,Σ,P) 是一个概率空间, 若事件 $A,B\in\Sigma$ 且满足 $P(AB)=P(A)P(B)$, 则称 **事件 A 与 B 是相互独立的**, 简称 **独立**.

根据定义可知任何事件与不可能事件 (或必然事件) 是相互独立的. 设两事件 A 和 B 是相互独立的, 且满足 $P(A)P(B) > 0$, 则有

$$P(AB) = P(A)P(B) \Leftrightarrow P(B|A) = P(B) \Leftrightarrow P(A|B) = P(A).$$

性质 2.4 若事件 A 与 B 相互独立, 则 A 与 $\bar{B}$, $\bar{A}$ 与 B, $\bar{A}$ 与 $\bar{B}$ 都互相独立.

证明 根据事件差公式 $P(A-B) = P(A) - P(AB)$ 有

$$P(A\bar{B}) = P(A-AB) = P(A) - P(AB) = P(A) - P(A)P(B) = P(A)(1-P(B)) = P(A)P(\bar{B}).$$

同理可证 $P(\bar{A}B) = P(\bar{A})P(B)$. 利用容斥原理有

$$\begin{aligned} P(\bar{A}\bar{B}) &= 1 - P(A \cup B) = 1 - P(A) - P(B) + P(AB) \\ &= 1 - P(A) - P(B) + P(A)P(B) = (1-P(A))(1-P(B)) = P(\bar{A})P(\bar{B}), \end{aligned}$$

从而完成证明.

独立与互斥之间的关系: 若事件 A 和 B 是独立的, 有 $P(AB) = P(A)P(B)$, 独立性与概率相关, 反映事件的概率属性; 若事件 A 和 B 是互斥的, 有 $AB = \emptyset$, 互斥性与事件的运算关系相关, 与概率无关, 因此独立性与互不相容性反映事件不同的属性.

如何判断事件的独立性? 可以先计算概率, 然后根据定义直接进行判断.

例 2.11 从一副不包括大王和小王的扑克中随机抽出一张, 用事件 A 表示抽到 10, 事件 B 表示抽到黑色的扑克, 讨论事件 A 与 B 的独立性?

解 不包括大王和小王的一副扑克共 52张, 黑色扑克 26 张, 4 张 10, 根据古典概型有

$$P(A) = 4/52 = 1/13, \quad P(B) = 1/2 \quad 和 \quad P(AB) = 1/26.$$

由此可得 $P(AB) = P(A)P(B)$, 根据定义可知事件 A 和 B 是相互独立的.

也可以根据实际问题判断事件的独立性, 例如

- 两人独立射击训练且互不影响, 因此两人中靶的事件相互独立;
- 有放回的产品抽样中第一次抽样结果和第二次抽样结果相互独立;
- 机器学习的经典假设是训练数据独立同分布采样.

针对条件概率, 可以考虑的条件独立性, 即在一定条件下两事件的独立性.

定义 2.4 设 (Ω, Σ, P) 是一个概率空间, 事件 $C \in \Sigma$ 有 $P(C) > 0$ 成立, 若事件 $A, B \in \Sigma$ 满足

$$P(AB|C) = P(A|C)P(B|C) \qquad \text{或} \qquad P(A|BC) = P(A|C),$$

则称事件 A 和 B 在 C 发生的情况下是 **条件独立的** (conditional independent).

下面给出一个关于条件独立性的例子:

例 2.12 盒子中有 $k+1$ 枚不均匀的硬币, 分别标号为 $0, 1, 2, \cdots, k$, 投掷第 i 号硬币正面向上的概率为 i/k. 现从箱子中任意取出一枚硬币并任意投掷多次, 若前 n 次正面向上, 求第 $n+1$ 次正面向上的概率.

解 用 A 表示第 $n+1$ 次投掷正面向上的事件, 用 B 表示前 n 次正面向上的事件, 用 C_i 表示从盒子中取出第 i 号硬币的事件. 根据条件概率的定义可知

$$P(A|B) = P(AB)/P(B) \ .$$

根据全概率公式和条件独立性有

$$P(AB) = \sum_{i=0}^{k} P(C_i)P(AB|C_i) = \sum_{i=0}^{k} P(C_i)P(A|C_i)P(B|C_i) = \frac{1}{k+1}\sum_{i=0}^{k}\frac{i^{n+1}}{k^{n+1}} \ ,$$

以及

$$P(B) = \sum_{i=0}^{k} P(C_i)P(B|C_i) = \frac{1}{k+1}\sum_{i=0}^{k}\frac{i^n}{k^n} \ ,$$

由此可得

$$P(A|B) = \frac{\sum_{i=0}^{k}(i/k)^{n+1}}{\sum_{i=0}^{k}(i/k)^n} \ .$$

当 k 非常大或 $k \to +\infty$ 时可利用积分近似

$$\frac{1}{k}\sum_{i=1}^{k}(i/k)^n \approx \int_0^1 x^n dx = \frac{1}{n+1} \quad \text{和} \quad \frac{1}{k}\sum_{i=1}^{k}(i/k)^{n+1} \approx \int_0^1 x^{n+1} dx = \frac{1}{n+2},$$

此时有 $P(A|B) \approx (n+1)/(n+2)$.

2.3.2 多个事件的独立性

定义 2.5 设 (Ω, Σ, P) 是一个概率空间, 若事件 $A, B, C \in \Sigma$ 满足

- 三事件两两独立, 即 $P(AB) = P(A)P(B)$, $P(AC) = P(A)P(C)$ 和 $P(BC) = P(B)P(C)$,
- $P(ABC) = P(A)P(B)P(C)$,

则称事件 A, B, C 是 **相互独立的**.