

第2章 条件概率与独立性

前面关于事件概率的讨论都是在整个样本空间上进行, 不需考虑其它条件或限制因素. 然而在很多实际问题中, 我们往往关心随机事件在一定附加信息 (条件) 下发生的概率, 即条件概率. 它是概率论中一个非常重要且实用的概念. 可以帮助我们更好地分析和理解复杂的随机事件, 同时也有助于简化复杂事件概率的计算.

2.1 条件概率

2.1.1 条件概率

首先来看一个例子, 随意投掷一枚骰子观察点数, 其样本空间 $\Omega = \{1, 2, \dots, 6\}$. 用 A 表示观察到奇数点的事件, 则事件 $A = \{1, 3, 5\}$, 根据古典概型有 $P(A) = 1/2$. 用 B 表示观察到 3 点的事件, 根据古典概型有 $P(B) = 1/6$.

现在考虑在事件 A 发生的情况下事件 B 发生的概率, 记为 $P(B|A)$. 由于 $A = \{1, 3, 5\}$ 且每种结果等可能发生, 由此可得条件概率

$$P(B|A) = 1/3 > P(B).$$

用事件 C 表示观察到 2 点, 根据古典概型有 $P(C) = 1/6$, 但在事件 A 发生的情况下事件 C 不可能发生, 因此条件概率

$$P(C|A) = 0 < P(C).$$

由此可知一个随机事件发生的概率可能随着条件的改变而改变, 同时通过观察可以发现

$$P(B|A) = 1/3 = \frac{P(AB)}{P(A)} \quad \text{和} \quad P(C|A) = 0 = \frac{P(AC)}{P(A)}.$$

针对一般情形, 我们将上述关系作为条件概率的定义.

定义 2.1 设 (Ω, Σ, P) 是一个概率空间, 随机事件 $A \in \Sigma$ 且 $P(A) > 0$. 对任意事件 $B \in \Sigma$, 称

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

为事件 A 发生的条件下事件 B 发生的概率, 简称 **条件概率** (conditional probability).

对任意事件 $A \in \Sigma$ 有 $P(A) = P(A\Omega)/P(\Omega) = P(A|\Omega)$ 成立, 因而任何随机事件的概率可以看作必然事件下的条件概率. 根据条件概率的定义有 $P(AB) = P(A)P(B|A)$. 在本书后续章节中, 若出现条件概率 $P(B|A)$, 一般都默认 $P(A) > 0$.

根据条件概率的定义, 不难验证条件概率 $P(\cdot|A)$ 具有以下一些基本性质:

1° **非负性**: 对任意事件 B 有 $P(B|A) \geq 0$;

2° **规范性**: 对样本空间 Ω 有 $P(\Omega|A) = 1$;

3° **可列可加性**: 若 $B_1, B_2, \dots, B_n, \dots$ 是可列无穷个互不相容的事件, 即 $B_i B_j = \emptyset$ ($i \neq j$), 有

$$P(B_1 \cup B_2 \cup \dots \cup B_n \cup \dots | A) = P(B_1|A) + P(B_2|A) + \dots + P(B_n|A) + \dots.$$

从概率和条件概率的定义可知 $P(B|A) = P(AB)/P(A) \geq 0$ 以及 $P(\Omega|A) = P(A\Omega)/P(A) = 1$, 由此验证公理 1° 和公理 2°. 若可列个事件 $B_1, B_2, \dots, B_n, \dots$ 是两两互不相容的, 则可列个事件 $AB_1, AB_2, \dots, AB_n, \dots$ 也是两两互不相容的, 根据分配律有

$$P\left(\bigcup_{i=1}^{\infty} B_i \middle| A\right) = \frac{P(A(\bigcup_{i=1}^{\infty} B_i))}{P(A)} = \frac{P(\bigcup_{i=1}^{\infty} AB_i)}{P(A)} = \sum_{i=1}^{\infty} \frac{P(AB_i)}{P(A)} = \sum_{i=1}^{\infty} P(B_i|A)$$

由此可知公理 3° 成立. 由于条件概率满足概率的三条公理, 因此条件概率 $P(\cdot|A)$ 仍然是一种概率.

性质 2.1 (容斥原理) 对随机事件 A, B_1 和 B_2 且满足 $P(A) > 0$, 有

$$P(B_1 \cup B_2 | A) = P(B_1 | A) + P(B_2 | A) - P(B_1 B_2 | A).$$

证明 由条件概率的定义有

$$P(B_1 \cup B_2 | A) = P((B_1 \cup B_2) \cap A) / P(A).$$

再根据随机事件的分配律和容斥原理有

$$P((B_1 \cup B_2) \cap A) = P(AB_1 \cup AB_2) = P(AB_1) + P(AB_2) - P(AB_1 B_2),$$

上式两边同时除以 $P(A)$ 即可完成证明.

性质 2.2 对随机事件 A 和 B 且满足 $P(A) > 0$, 有 $P(\bar{B}|A) = 1 - P(B|A)$.

证明 根据容斥原理有

$$1 = P(\Omega|A) = P(B \cup \bar{B}|A) = P(B|A) + P(\bar{B}|A) - P(B\bar{B}|A)$$

再根据事件 B 和 $\bar{B}$ 互不相容有 $P(B\bar{B}|A) = 0$, 从而完成证明.

事件 A 发生的条件下事件 B 发生的概率, 可以将 A 看作新的样本空间、而忽略以前的样本空间 Ω , 由此可以发现: **条件概率的本质是缩小了有效的样本空间**. 这也为计算条件概率提供了一种方法: 空间缩减法, 在新的样本空间 A 下考虑事件 B 发生的概率.

例 2.1 盒子中有 4 个不同的产品, 其中 3 个一等品, 1 个二等品. 从盒子中不放回随机取两次产品. 用 A 表示第一次拿到一等品的事件, B 表示第二次取到一等品的事件, 求条件概率 $P(B|A)$.

解 将盒子中 3 个一等产品分别编号为 1, 2, 3, 二等品编号 4. 用 i 和 j 分别表示第一、二次抽取的产品的编号, 由此可得

$$\begin{aligned}\Omega &= \{(i, j): i \neq j, i, j \in [4]\}, & A &= \{(i, j): i \neq j, i \neq 4\}, \\ B &= \{(i, j): i \neq j, j \neq 4\}, & AB &= \{(i, j): i \neq j, i, j \in [3]\}.\end{aligned}$$

计算可得 $|\Omega| = 12$, $|A| = 9$, $|B| = 9$ 以及 $AB = 6$. 根据古典概型有

$$P(A) = \frac{|A|}{|\Omega|} = \frac{3}{4}, \quad P(B|A) = \frac{P(AB)}{P(A)} = \frac{1/2}{3/4} = \frac{2}{3}.$$

也可以采用 **样本空间缩减法** 来求解此问题: 当事件 A 发生后, 剩下 2 只一等品, 1 只二等品, 因此直接得到 $P(B|A) = 2/3$.

例 2.2 箱子中有 a 个红球和 b 个白球, 依次任意无放回地取出 n 个球 ($n \leq a + b$), 其中包括 k 个白球 ($k \leq b$), 在此情形下求第一次取出白球的概率.

解 用 A 表示第一次取出白球的事件, 用 B 表示依次取出 n 个球中包括 k 个白球的事件. 根据题意和超几何分布有

$$P(A) = \frac{b}{a+b}, \quad P(B) = \frac{\binom{a}{n-k} \binom{b}{k}}{\binom{a+b}{n}}, \quad P(B|A) = \frac{\binom{a}{n-k} \binom{b-1}{k-1}}{\binom{a+b-1}{n-1}}.$$

所求条件概率

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B|A)}{P(B)} = \frac{k}{n}.$$

也可以采用 **样本空间缩减法** 来求解此问题: 在事件 B 发生的情况下, 即选中的 n 个球中有 k 个白球, 由于任何一球被第一次选中的可能性一样, 因此事件 A 发生的概率为 k/n .

2.1.2 乘法公式

随机事件 A 和 B 满足 $P(A) > 0, P(B) > 0$, 根据条件概率的定义可知

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

将上式进一步推广, 根据条件概率的定义有下面的乘法公式:

定理 2.1 设 (Ω, Σ, P) 是一个概率空间, 若随机事件 $A_1, A_2, \dots, A_n \in \Sigma$, 且满足条件 $P(A_1 A_2 \cdots A_{n-1}) > 0$, 则有

$$P(A_1 A_2 \cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 A_2) \cdots P(A_n|A_1 A_2 \cdots A_{n-1}).$$

例 2.3 假设一批灯泡有 100 只, 其中有次品 10 只, 其余为正品. 不放回抽取地每次抽取一只, 求第三次才是正品的概率.

解 用 A_i 表示第 i 次抽到正品的事件 ($i \in [3]$), 事件 B 表示第 3 次才抽到的正品, 则有 $B = \bar{A}_1 \bar{A}_2 A_3$. 根据乘法公式有

$$P(B) = P(\bar{A}_1 \bar{A}_2 A_3) = P(\bar{A}_1)P(\bar{A}_2|\bar{A}_1)P(A_3|\bar{A}_1 \bar{A}_2) = \frac{10}{100} \times \frac{9}{99} \times \frac{90}{98} = \frac{9}{1078}.$$

例 2.4 设 n 把钥匙中只有一把能打开门. 不放回随机取出一把开门, 求第 k 次打开门的概率.

解 用 A_i 表示第 i 次没有打开门的事件, 则第 k 次打开门的事件可表示为 $A_1 A_2 \cdots A_{k-1} \bar{A}_k$, 根据乘法公式有

$$\begin{aligned} P(A_1 A_2 \cdots A_{k-1} \bar{A}_k) &= P(A_1)P(A_2|A_1)P(A_3|A_1 A_2) \cdots P(A_{k-1}|A_1 \cdots A_{k-2})P(\bar{A}_k|A_1 A_2 \cdots A_{k-1}) \\ &= \frac{n-1}{n} \times \frac{n-2}{n-1} \times \cdots \times \frac{n-(k-1)}{n-(k-2)} \times \frac{1}{n-(k-1)} = \frac{1}{n} \end{aligned}$$

也可以根据 **抽签原理** 来求解该问题: 第 k 次打开门的概率与 k 无关, 每次打开门的概率相同, 共 n 把钥匙, 因此第 k 次打开门的概率为 $1/n$.

例 2.5 假设有 n 对夫妻参加活动, 被随机分成 n 组, 每组一男一女, 求 n 对夫妻恰好两两被分到一组的概率.

解 用 A_i 表示第 i 对夫妻被分到同一组的事件, 则 n 对夫妻恰好两两被分到一组的事件可表示为 $A_1 A_2 \cdots A_n$. 根据乘法公式有

$$\begin{aligned} P(A_1 A_2 \cdots A_n) &= P(A_1)P(A_2|A_1)P(A_3|A_1 A_2) \cdots P(A_n|A_1 A_2 \cdots A_{n-1}) \\ &= \frac{1}{n} \times \frac{1}{n-1} \times \cdots \times \frac{1}{1} = \frac{1}{n!}. \end{aligned}$$

例 2.6 第一个箱子里有 n 个不同的白球, 第二个箱子里有 m 个不同的红球, 从第一个箱子任意取走一球, 再从第二个箱子里任意取走一球放入第一个箱子, 依次进行, 直至第一、第二个箱子都为空, 求第一个箱子最后一次取走的球是白球的概率.

解 假设第一个箱子里的白球分别标号为 $1, 2, \cdots, n$, 用 A_i 表示第一个箱子最后取走的是第 i 号白球的事件. 由此可知事件 $A_1, A_2, \cdots, A_n$ 是两两互不相容的, 且第一个箱子最后一次取走的球是白球的事件可表示为 $A_1 \cup A_2 \cup \cdots \cup A_n$, 根据事件的对称性可得其概率为

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^n P(A_i) = nP(A_1).$$

若事件 A_1 发生, 则从第一个箱子中取走的 $m+n-1$ 个球均不是第 1 号白球, 用事件 B_j 表示第 j 次从第一个箱子里取走的球不是第 1 号白球, 即 $A_1 = B_1 B_2 \cdots B_{m+n-1}$. 根据乘法公式有

$$\begin{aligned} P(A_1) &= P(B_1)P(B_2|B_1) \cdots P(B_m|B_1 B_2 \cdots B_{m-1}) \times P(B_{m+1} B_{m+2} \cdots B_{m+n-1} | B_1 B_2 \cdots B_m) \\ &= \left(1 - \frac{1}{n}\right)^m P(B_{m+1} B_{m+2} \cdots B_{m+n-1} | B_1 B_2 \cdots B_m) = \left(1 - \frac{1}{n}\right)^m \times \frac{1}{n}. \end{aligned}$$

由此可知第一个箱子最后一次取走的球是白球的概率为 $(1 - 1/n)^m$.

例 2.7 假设箱子里有 m 个红球和 n 个白球, 现随机取出一球后放回, 并加入 c 个与取出球同色的球, 求前两次取出红球、后两次取出白球的概率.

解 用 A_i 表示第 i 次抽到红球的事件 ($i \in [2]$), 事件 B_i 表示第 i 次抽到白球的事件 ($i = 3, 4$), 我们有

$$\begin{aligned} P(A_1) &= \frac{m}{m+n}, & P(A_2|A_1) &= \frac{m+c}{m+n+c}, \\ P(B_1|A_1 A_2) &= \frac{n}{m+n+2c}, & P(B_2|A_1 A_2 B_1) &= \frac{n+c}{m+n+3c}. \end{aligned}$$

根据乘法公式有

$$P(A_1 A_2 B_1 B_2) = \frac{mn(m+c)(n+c)}{(m+n)(m+n+c)(m+n+2c)(m+n+3c)}$$

上述例子可用来作为疾病传染的粗略解释, 每取出一球代表疾病的一次传染, 每次传染将增加再传染的可能性.

2.2 全概率公式和贝叶斯公式

本节介绍概率计算中两个重要的公式: 全概率公式和贝叶斯公式.

2.2.1 全概率公式

全概率公式是概率论中最基本的公式之一, 将一个复杂事件的概率计算分解为若干简单事件的概率计算. 具体而言, 将一个复杂事件分解为若干不相容的简单事件之和, 通过分别计算简单事件的概率, 利用概率的可加性得到复杂事件的概率. 首先定义样本空间的一个分割.

定义 2.2 设 $A_1, A_2, \cdots, A_n$ 是样本空间 Ω 的一组事件, 若满足:

- i) 任意两个事件是互不相容性的, 即 $A_i \cap A_j = \emptyset$ ($i \neq j$);
- ii) 完备性 $\Omega = A_1 \cup A_2 \cup \cdots \cup A_n$,

则称事件 $A_1, A_2, \cdots, A_n$ 为样本空间 Ω 的一个 **分割**, 亦称 **完备事件组**.

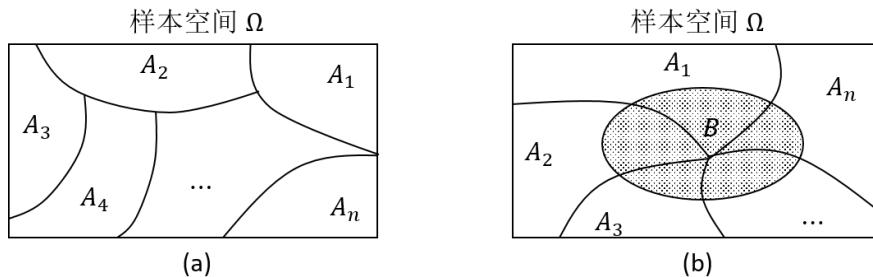


图 2.1 样本空间的分割与事件

如图 2.1(a) 所示, 若 $A_1, A_2, \dots, A_n$ 为样本空间 Ω 的一个分割, 则每次试验时在 $A_1, A_2, \dots, A_n$ 中有且仅有一个事件发生. 对任何事件 $A \subseteq \Omega$, 事件 A 与对立事件 $\bar{A}$ 构成样本空间 Ω 的一个分割.

基于样本空间的分割, 下面给出全概率公式:

定理 2.2 设事件 $A_1, A_2, \dots, A_n$ 是样本空间 Ω 的一个分割, 对任意事件 B 有

$$P(B) = \sum_{i=1}^n P(BA_i) = \sum_{i=1}^n P(A_i)P(B|A_i),$$

该公式被称为 **全概率公式** (Law of total probability).

证明 该定理的证明本质上是对加法和乘法事件的综合运用. 首先根据分配律有

$$B = B \cap \Omega = B \cap \left(\bigcup_{i=1}^n A_i \right) = \bigcup_{i=1}^n BA_i$$

由 $A_i \cap A_j = \emptyset$ 可得 $BA_i \cap BA_j = \emptyset$, 由概率的有限可列可加性有

$$P(B) = P\left(\bigcup_{i=1}^n BA_i\right) = \sum_{i=1}^n P(BA_i) = \sum_{i=1}^n P(A_i)P(B|A_i).$$

由于任意事件和其对立事件构成一个分割, 对任意概率非零的事件 A 和 B 有

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

如图 2.1(b), 还可以从另一个角度来理解全概率公式: 将事件 B 看作一个结果, 将事件 $A_1, A_2, \dots, A_n$ 看作产生该结果的若干原因, 针对不同原因事件 B 发生的概率 (即条件概率 $P(B|A_i)$) 各不相同, 而到底是哪一种原因具有随机性. 具体而言, 每一种原因发生的概率 $P(A_i)$ 是已知的, 以及每一种原因对结果 B 的影响 $P(B|A_i)$ 已知, 则可以计算结果 $P(B)$.

下面来看一些例子:

例 2.8 小明参加一次人工智能竞赛, 目前的排名不理想, 分析其原因: 方法不够新颖的概率为 50%, 通过设计新方法后取得理想排名的概率为 50%; 程度代码有误的概率为 30%, 通过纠正代码后

取得理想排名的概率为 60%; 数据不充分的概率为 20%, 通过采集更多数据后取得理想排名的概率为 80%. 求小明最后取得理想排名的概率.

解 用 B 表示小明最后取得理想排名的事件, 用 A_1, A_2, A_3 分别表示方法不够新颖、程度代码有误、数据不充分这三个事件, 根据题意有

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

小明最后取得理想排名的概率

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

例 2.9 随意抛 n 次硬币, 证明正面朝上的次数是偶数 (或奇数) 的概率为 $1/2$.

证明 用事件 A 表示前 $n-1$ 次抛硬币正面朝上的次数为偶数, 其对立事件 $\bar{A}$ 表示前 $n-1$ 次抛硬币朝上的次数为奇数, 事件 B 表示前 n 次硬币朝上的次数为偶数. 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. 若正面朝上的次数是偶数, 则随意抛 n 次硬币中正面朝上的次数为偶数分别有 $\{0, 2, 4, \dots, 2k\}$ ($2k \leq n$), 根据概率公式直接计算有

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

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

例 2.10 假设有 n 个箱子, 每个箱子里有 a 只白球和 b 只红球, 现从第一个箱子取出一个球放入第二个箱子, 第二个箱子取出一个球放入第三个箱子, 依次类推, 求从最后一个箱子取出一球是红球的概率.

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

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

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, 最后一个箱子中取出一球是红球的概率为 $b/(a+b)$.

2.2.2 贝叶斯公式

贝叶斯公式也是概率论中最基本的公式之一, 在结果发生的情况下探讨是由何种原因导致结果. 具体而言, 假设有 $A_1, A_2, \dots, A_n$ 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 $A_1, A_2, \dots, A_n$ 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 $1 \leq i \leq n$ 有

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

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 $A_1, A_2, \dots, A_n$ 看作事件 B 发生的“原因”, 而事件 B 是伴随着原因 $A_1, A_2, \dots, A_n$ 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, 此时利用贝叶斯公式探讨是由某原因 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 2.11 小明参加一次人工智能竞赛, 目前的排名不理想, 分析其原因: 方法不够新颖的概率为 50%, 通过设计新方法后取得理想排名的概率为 50%; 程度代码有误的概率为 30%, 通过纠正代码后取得理想排名的概率为 60%; 数据不充分的概率为 20%, 通过采集更多数据后取得理想排名的概率为 80%. 因为时间有限, 小明只能选择三种方案 (设计新方法、纠正代码、采集更多数据) 中一种, 想要取得理想的排名, 小明应该选择哪一种方案.

解 用 B 表示小明最后取得理想排名的事件, 用 A_1, A_2, A_3 分别表示方法不够新颖、程度代码有误、数据不充分这三个事件, 根据题意有

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

小明最后取得理想排名的概率

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

根据贝叶斯公式有 $P(A_1|B) = P(A_1)P(B|A_1)/P(B) = 25/59$, 同理可得 $P(A_2|B) = 18/59$ 和 $P(A_3|B) = 16/59$. 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 2.12 (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

$$P(D|A) = 1/2 \quad P(D|B) = 0 \quad P(D|C) = 1.$$

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.

贝叶斯公式提出了重要的推理逻辑, 在概率统计以及日常生活中存在多方面的应用. 假定 $A_1, A_2, \dots, A_n$ 是导致结果事件 B 的“原因”, 概率 $P(A_i)$ 被称为 **先验概率** (prior probability), 反映了各种“原因”的可能性大小, 一般都是根据先前的经验总结而成. 若现在试验产生了事件 B , 这个信息有助于探讨事件发生的“原因”, 条件概率 $\Pr(A_i|B)$ 被称 **后验概率** (posterior probability), 反映了试验之后对各种“原因”发生可能性的新知识.

例如, 医生为诊断病人患了疾病 $A_1, A_2, \dots$ 中哪一种疾病, 可以对病人进行检查, 确定某个指标 B (如血糖、血脂、血钙等), 从而帮助诊断, 此时可以采用贝叶斯公式来计算相关概率. 根据以往的数据资料确定先验概率 $P(A_i)$, 即人们患各种疾病的可能性; 在通过医学知识确定概率 $P(B|A_i)$, 最后通过贝叶斯公式计算后验概率 $P(A_i|B)$. 在实际应用中, 可能检验多个指标 B , 综合所有的后验概率进行诊断. 在自动诊断和辅助诊断的专家系统中, 这种方法非常实用.

贝叶斯公式使用中最存在争议之处在于先验的选取, 在很多实际应用中往往都根据以往的数据而得出的, 符合概率的频率解释, 但需要以往大量的历史数据, 在实际应用中通常难以满足. 其次, 在很多应用中先验概率可能由某一种主观的方式给出, 例如对未来宏观经济形势 (或对某人诚信度) 的判断, 这种将概率解释为信任程度的做法明显带有主观性, 通常被称为 **主观概率**.

伊索寓言“孩子与狼”讲一个小孩每天到山上放羊, 山里有狼出没, 第一天他在山上喊“狼来了! 狼来了!”, 山下的村民们闻声便去打狼, 到了山上发现没有狼; 第二天仍是如此; 第三天狼真来了, 可无论小孩怎么喊叫, 也没有人来救他, 因为前二次他说了谎话, 人们不再相信他了. 我们可以将这个寓言抽象为一个主观概率的例子, 并利用贝叶斯公式来分析这个寓言中村民们的心理活动.

例 2.13 假设村民们对这个小孩的印象一般, 认为小孩说谎话和说真话的概率相同, 均为 $1/2$. 假设说谎话的小孩喊狼来了时狼真来的概率为 $1/3$, 而说真话的小孩喊狼来了时狼真来的概率为 $3/4$. 若第一天、第二天上山均没有发现狼, 请分析村民们的心理活动.

解 用 B_1 和 B_2 分别表示第一天和第二天狼来了的事件, 用 A_1 表示小孩第一天说谎话的事件, 用 A_2 表示在第一天狼没有的情况下小孩第二天说谎话的事件, 根据题意可知

$$P(A_1) = P(\overline{A_1}) = 1/2, P(B_1|A_1) = 1/3, P(B_1|\overline{A_1}) = 3/4, P(B_2|A_2) = 1/3, P(B_2|\overline{A_2}) = 3/4.$$

第一天村民上山打狼但没有发现狼, 根据贝叶斯公式可知村民们对说谎话小孩的认识发生了改变, 体现在

$$P(A_2) = P(A_1|\overline{B_1}) = \frac{P(\overline{B_1}|A_1)P(A_1)}{P(\overline{B_1}|A_1)P(A_1) + P(\overline{B_1}|\overline{A_1})P(\overline{A_1})} = \frac{8}{11} \approx 0.7273, \quad P(\overline{A_2}) = \frac{3}{11}.$$

此时, 村民对这个小孩说谎话的概率从 50% 调整到 72.72%.

第二天村民上山打狼还是没有发现狼, 根据贝叶斯公式可知村民们对说谎话小孩的认识又发生了改变, 体现在

$$P(A_2|\overline{B_2}) = \frac{P(\overline{B_2}|A_2)P(A_2)}{P(\overline{B_2}|A_2)P(A_2) + P(\overline{B_2}|\overline{A_2})P(\overline{A_2})} = \frac{64}{73} \approx 0.8767.$$

此时, 村民对这个小孩说谎话的概率从 72.72% 调整到 87.67%.

这表明村民们经过两次上当, 对这个小孩说谎话的概率从 50% 上升到 87.67%, 给村民留下这种印象, 他们听到第三次呼叫时不会再上山打狼.

2.3 事件独立性

前面的例子表明, 在事件 A 发生的条件下事件 B 发生的条件概率 $P(B|A)$, 通常不等于事件 B 发生的概率 $P(B)$ (无任何附加条件), 即 $P(B|A) \neq P(B)$, 也就是说“事件 A 发生通常会改变事件 B 发生的可能性”. 然而在有些特殊情形下, 事件 A 的发生对事件 B 的发生可能没有任何影响, 这就是本节所研究的事件独立性.

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 2.3 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.

如何判断事件的独立性? 根据定义直接计算进行判断:

例 2.14 从一副扑克 (不含大王、小王) 中随机抽取一张扑克, 用事件 A 表示抽到 10, 事件 B 表示抽到黑色的扑克. 事件 A 与 B 是否独立?

解 根据问题可知一副扑克 (不含大王、小王) 52 张, 黑色扑克 26 张, 4 张 10, 根据古典概型有

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

由此可得 $P(AB) = 2/52 = 1/26 = P(A)P(B)$, 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

- 两人独立射击打靶、且互不影响, 因此两人中靶的事件相互独立;
- 从 n 件产品中随机抽取两件, 事件 A_i 表示第 i 件是合格品. 若有放回抽取则事件 A_1 与 A_2 相互独立; 若不放回则不独立;
- 机器学习的经典假设是训练数据独立同分布采样.

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性质.

类似于条件概率, 可以定义概率论中的条件独立性, 即在一定条件下两事件是相互独立的.

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

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

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 2.15 假设一个箱子中有 $k+1$ 枚不均匀的硬币, 投掷第 i 枚硬币时正面向上的概率为 i/k ($i = 0, 1, 2, \dots, k$). 现从箱子中任意取出一枚硬币、并任意重复投掷多次, 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 次投掷都正面向上的事件, 用 C_i 表示从箱子中取出第 i 枚硬币的事件 ($i = 0, 1, 2, \dots, k$). 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 $k \rightarrow +\infty$ 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 $A, B, C \in \Sigma$ 且满足

- 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**.

根据定义可知若事件 A, B, C 是相互独立的, 则事件 A, B, C 是两两相互独立的; 但反之不一定成立, 还需满足 $P(ABC) = P(A)P(B)P(C)$. 下面给出一个简单的例子说明: 三事件的两两独立并不能得出三事件相互独立.

[Bernstein 反例] 一个均匀的正四面体, 第一面是红色, 第二面是白色, 第三面是黑色, 第四面同时有红、白、黑三种颜色. 随意投掷一次该四面体, 用 A, B, C 分别表示红色、白色、黑色朝下的事件, 因为有一面同时包含三种颜色, 有

$$P(A) = P(B) = P(C) = 1/2 \quad \text{和} \quad P(AB) = P(BC) = P(AC) = 1/4,$$

由此可得事件 A, B, C 两两独立. 但由于

$$P(ABC) = 1/4 \neq 1/8 = P(A)P(B)P(C),$$

由此可知 A, B, C 不是相互独立的.

定义 2.6 设 (Ω, Σ, P) 是一个概率空间, 若事件 $A_1, A_2, \dots, A_n \in \Sigma$ 中任意 k 个事件是相互独立的 ($k \geq 2$), 即满足

- 对任意 $1 \leq i_1 < i_2 \leq n$ 有 $P(A_{i_1}A_{i_2}) = P(A_{i_1})P(A_{i_2})$ 成立;
- 对任意 $1 \leq i_1 < i_2 < i_3 \leq n$ 有 $P(A_{i_1}A_{i_2}A_{i_3}) = P(A_{i_1})P(A_{i_2})P(A_{i_3})$ 成立;
- $\dots \dots$
- $P(A_1A_2 \cdots A_n) = P(A_1)P(A_2) \cdots P(A_n)$,

则称事件 $A_1, A_2, \dots, A_n$ 是 **相互独立的**.

根据定义可知, n 个事件的相互独立性应满足 $\binom{n}{2} + \binom{n}{3} + \cdots + \binom{n}{n} = 2^n - n - 1$ 个等式, 同样 n 个事件的相互独立性与两两独立性是不同的概念. 类似地可以定义多个事件的条件独立性. 下面看一个关于独立性的例子.

例 2.16 三人独立破译一份密码, 每人单独能破译的概率分别为 $1/5, 1/3, 1/4$, 问三人中至少有一人能破译密码的概率.

解 用事件 A_i 表示第 i 个人破译密码 ($i \in [3]$), 根据题意有

$$P(A_1) = 1/5, \quad P(A_2) = 1/3, \quad P(A_3) = 1/4.$$

根据容斥原理和独立性, 三人中至少有一人能破译密码的概率为

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1A_2) - P(A_1A_3) - P(A_2A_3) + P(A_1A_2A_3) = 3/5.$$

也可以根据对偶性和独立性来求解该问题, 三人中至少有一人能破译密码的概率为

$$P(A_1 \cup A_2 \cup A_3) = 1 - P(\bar{A}_1 \bar{A}_2 \bar{A}_3) = 1 - P(\bar{A}_1)P(\bar{A}_2)P(\bar{A}_3) = 1 - \frac{4}{5} \cdot \frac{2}{3} \cdot \frac{3}{4} = 3/5.$$

从上例可知: 尽管每个人能破译密码的概率都不大于 $1/3$, 但三人独立进行破译, 则至少有一人破译密码的概率则为 $3/5$, 由此提高了破译密码的概率. 我们可以将类似问题推广到更一般的情况.

若事件 $A_1, A_2, \dots, A_n$ 相互独立, 发生的概率分别为 $p_1, p_2, \dots, p_n$, 则事件 $A_1, A_2, \dots, A_n$ 中至少有一事件发生的概率为

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = 1 - P(\bar{A}_1 \bar{A}_2 \dots \bar{A}_n) = 1 - (1 - p_1)(1 - p_2) \dots (1 - p_n);$$

此外, 事件 $A_1, A_2, \dots, A_n$ 中至少有一事件不发生的概率为

$$P(\bar{A}_1 \cup \bar{A}_2 \cup \dots \cup \bar{A}_n) = 1 - P(A_1 A_2 \dots A_n) = 1 - p_1 p_2 \dots p_n.$$

由此可知: 尽管每个事件发生的概率 p_i 都非常小, 但若 n 非常大, 则 n 个相互独立的事件中“至少有一事件发生”或“至少有一事件不发生”的概率可能很大.

定义 2.7 (小概率原理) 若事件 A 在一次试验中发生的概率非常小, 但经过多次独立地重复试验, 事件 A 的发生是必然的, 称之为 **小概率原理**.

小概率原理可通过严格的数学证明得到: 若事件 $A_1, A_2, \dots, A_n, \dots$ 独立且每事件发生的概率 $P(A_i) = p > 0$ 非常小, 则有

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = 1 - P(\bar{A}_1 \bar{A}_2 \dots \bar{A}_n) = 1 - (1 - p)^n \rightarrow 1 \quad \text{当} \quad n \rightarrow \infty,$$

即独立重复多次的小概率事件亦可成立必然事件.

例 2.17 冷战时期美国的导弹精度 90%, 苏联的导弹精度 70%, 但苏联的导弹数量特别多, 导弹的数量能否弥补精度的不足?

解 假设每次独立发射 n 枚导弹, 用事件 A_i 表示第 i 枚导弹命中目标, 则 n 枚导弹击中目标的概率为

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = 1 - (1 - 0.7)^n \geq 0.9 \quad \Rightarrow \quad n \geq 2,$$

因此每次独立发射 2 枚导弹, 击中目标的概率高于 90%.

例 2.18 假设市场上有 m 种不同类型的邮票, 一位集邮爱好者收集第 i 种邮票的概率为 p_i , 且 $p_1 + p_2 + \dots + p_m = 1$. 假设每次集邮都是独立同分布的, 若现已收集到 n 张邮票, 用 A_i 表示至少收集到第 i 种类型邮票的事件, 求概率 $P(A_i)$, $P(A_i \cup A_j)$ 以及 $P(A_i | A_j)$ ($i \neq j$).

解 根据题意有

$$P(A_i) = 1 - P(\overline{A_i}) = 1 - P(\text{收集的 } n \text{ 张邮票中没有第 } i \text{ 种类型邮票}) = 1 - (1 - p_i)^n.$$

同理可得

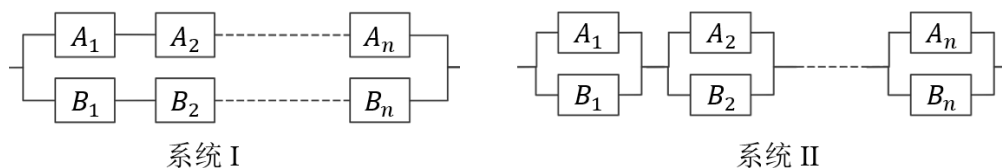
$$P(A_i \cup A_j) = 1 - P(\overline{A_i} \cap \overline{A_j}) = 1 - (1 - p_i - p_j)^n.$$

利用容斥原理和条件概率的定义有

$$\begin{aligned} P(A_i \cup A_j) &= \frac{P(A_i A_j)}{P(A_j)} = \frac{P(A_i) + P(A_j) - P(A_i \cup A_j)}{P(A_j)} \\ &= \frac{1 - (1 - p_i)^n - (1 - p_j)^n + (1 - p_i - p_j)^n}{1 - (1 - p_j)^n}. \end{aligned}$$

为保证系统的可靠性, 近代电子系统通常由多个独立的元件构成, 一个元件能正常工作的概率称为这个元件的可靠性. 由元件组成的系统能正常工作的概率称为系统的可靠性.

例 2.19 设构成系统的每个元件的可靠性均为 p ($0 < p < 1$), 且各元件是否正常工作是相互独立的. 设有 $2n$ 个元件按下图所示, 两种不同连接方式构成两个不同的系统, 比较这两种系统的可靠性大小.



解 用事件 A_i 和 B_i 表示图中所对应的元件正常工作 ($i = 1, 2, \dots, n$). 可以发现系统 I 有两条通路, 它能正常工作当且仅当两条通路至少有一条能正常工作, 而每一条通路能正常工作当且仅当它的每个元件能正常工作, 因此有系统 I 的可靠性为

$$\begin{aligned} &P((A_1 A_2 \cdots A_n) \cup (B_1 B_2 \cdots B_n)) \\ &= P(A_1 A_2 \cdots A_n) + P(B_1 B_2 \cdots B_n) - P(A_1 A_2 \cdots A_n B_1 B_2 \cdots B_n) = 2p^n - p^{2n} = p^n(2 - p^n). \end{aligned}$$

系统 II 由 n 对并联元件 $\{A_i, B_i\}$ 组成, 它能正常工作当且仅当每对并联元件组能够正常工作, 因此系统 II 的可靠性为

$$P\left(\bigcap_{i=1}^n (A_i \cup B_i)\right) = \prod_{i=1}^n P(A_i \cup B_i) = (2p - p^2)^n = p^n(2 - p)^n.$$

利用数学归纳法可证明当 $n \geq 2$ 时有 $(2 - p)^n > 2 - p^n$ 成立, 由此可知系统 II 的可靠性更好.

2.3.3 Borel-Cantelli 引理*

Borel-Cantelli 引理常常被用来计算事件的概率为 0 或 1, 首先介绍一个有用的引理:

引理 2.1 若数列 $\{p_i\}_{i=1}^n$ 满足 $p_i \in [0, 1]$ 和 $\sum_{i=1}^{\infty} p_i = +\infty$, 则有 $\prod_{i=1}^{\infty} (1 - p_i) = 0$.

证明 对任意 $x \in [0, 1]$, 有 $\ln(1 - x) \leq -x$, 于是得到

$$\ln \prod_{i=1}^{\infty} (1 - p_i) \leq \ln \prod_{i=1}^n (1 - p_i) = \sum_{i=1}^n \ln(1 - p_i) \leq \sum_{i=1}^n -p_i.$$

分别对上式两边取极限 $n \rightarrow +\infty$ 有 $\ln \prod_{i=1}^{\infty} (1 - p_i) = -\infty$, 由此完成证明.

根据上述引理, 我们可以证明如下定理

定理 2.4 (Borel-Cantelli 引理) 设 (Ω, Σ, P) 是一个概率空间, 以及事件系列 $A_i \in \Sigma$, 令事件 $A = \bigcap_{n=1}^{+\infty} \bigcup_{i=n}^{+\infty} A_i$, 则有

- 若 $\sum_{i=1}^{\infty} P(A_i) < +\infty$ 则有 $P(A) = 0$;
- 若 $\sum_{i=1}^{\infty} P(A_i) = +\infty$ 且事件 $\{A_i\}$ 相互独立, 则有 $P(A) = 1$.

该定理考虑事件序列 $\{A_i\}_{i=1}^{+\infty}$ 中属于无限多 A_i 的基本事件的概率和. 在定理第二不妨中事件 A_i 之间相互独立

证明 根据无穷级数 $\sum_{i=1}^{\infty} P(A_i) < +\infty$ 收敛的性质可知 $\lim_{n \rightarrow \infty} \sum_{i=n}^{\infty} P(A_i) = 0$. 根据题意可知 $A \subseteq \bigcup_{i=n}^{+\infty} A_i$, 利用 Union bounds 有

$$P(A) \leq P\left(\bigcup_{i=n}^{+\infty} A_i\right) \leq \sum_{i=n}^{+\infty} P(A_i),$$

上式两边同时对 $n \rightarrow +\infty$ 取极限证明 $P(A) = 0$.

针对第二个问题, 不妨设 $B_n = \bigcup_{i=n}^{+\infty} A_i$, 由此可知 $A = \bigcap_{n=1}^{+\infty} B_n$. 给定任意正整数 $m > n \geq 1$, 利用德摩根律和独立性假设有

$$P(\overline{B_n}) = P\left(\bigcap_{i=n}^{\infty} \overline{A_i}\right) = \lim_{m \rightarrow +\infty} P\left(\bigcap_{i=n}^m \overline{A_i}\right) = \lim_{m \rightarrow +\infty} \prod_{i=n}^m P(\overline{A_i}) = \prod_{i=n}^{+\infty} (1 - P(A_i))$$

根据引理 2.1 可得 $P(\overline{B_n}) = 0$, 结合德摩根律进一步有

$$P(\overline{A}) = P\left(\bigcup_{n=1}^{+\infty} \overline{B_n}\right) \leq \sum_{n=1}^{+\infty} P(\overline{B_n}) = 0,$$

由此完成证明.

2.4 案例分析

下面将利用本节知识来解决一些实际的问题, 值得注意的是贝叶斯公式在人工智能的决策任务中有诸多的应用, 例如朴素贝叶斯分类器等, 由于涉及到多维随机变量相关, 我们将在后面的章节中介绍贝叶斯公式的应用.

2.4.1 多项式相等

有两个较为复杂的多项式, 例如

$$\begin{aligned} F(x) &= (x+2)^7(x+3)^5 + (x+1)^{100} + (x+2)(x+3) + x^{20}, \\ G(x) &= (x+3)^{100} - (x+1)^{25}(x+2)^{30} + x^{20} + (x-2)(x-3) \cdots (x-100). \end{aligned}$$

是否存在一种方法验证 $F(x) \equiv G(x)$.

最容易想到的方法是将多项式全部展开, 合并同类项, 比较多项式每项的系数, 若相应的系数完全相同则有 $F(x) \equiv G(x)$. 但这种方法通常需要较高的计算时间开销, 当多项式较复杂时更加困难, 是否存在一种简单快捷的验证方法.

我们介绍一种利用随机性来求解该问题的简单方法: 不妨假设 $F(x)$ 和 $G(x)$ 的最高次 (或多项式的度) 不超过 d , 考虑从集合 $[100d] = \{1, 2, \dots, 100d\}$ 中等可能随意选取一个数 r , 然后计算 $F(r)$ 和 $G(r)$, 若 $F(r) \neq G(r)$ 则返回 $F(x) \not\equiv G(x)$; 否则返回 $F(x) \equiv G(x)$. 下面分析该方法的正确性:

- 若多项式 $F(x) \equiv G(x)$, 则该方法得到“正确”结果, 因为对任意 $r \in [100d]$ 都有 $F(r) = G(r)$.
- 若多项式 $F(x) \not\equiv G(x)$ 且 $F(r) \neq G(r)$, 则该方法也得到“正确”结果, 因为找到了一个 $r \in [100d]$ 使得 $F(r) \neq G(r)$ 成立.
- 若多项式 $F(x) \not\equiv G(x)$ 但 $F(r) = G(r)$, 则该方法得到“错误”结果. 当 $F(x) \not\equiv G(x)$ 时, 依然存在 $r \in [100d]$ 使得 $F(r) = G(r)$ 成立, 此时 r 是多项式 $F(x) - G(x) = 0$ 的一个实数根. 根据代数知识不超过 d 次多项式 $F(x) - G(x) = 0$ 至多有 d 个实数根, 而 r 从 $[100d]$ 中等可能随机选取, 因此有

$$P[F(r) = G(r)] \leq d/100d = 1/100.$$

利用独立性可以进一步提高方法返回“正确”的概率: 从集合 $[100d]$ 中独立地随意选取 k ($< d$) 个数 $r_1, r_2, \dots, r_k$. 若存在 r_i 使得 $F(r_i) \neq G(r_i)$ 成立, 则返回 $F(x) \not\equiv G(x)$, 否则返回 $F(x) \equiv G(x)$.

这里仅分析该方法返回“错误”结果发生的概率, 当 $F(x) \not\equiv G(x)$ 时出现 $F(r_1) = G(r_1), F(r_2) = G(r_2), \dots, F(r_k) = G(r_k)$ 的概率, 根据事件的独立性与前面的分析有

$$P\left(\bigcap_{i=1}^k \{F(r_i) = G(r_i)\}\right) = \prod_{i=1}^k P(F(r_i) = G(r_i)) \leq 1/100^k,$$

因此显著提高了方法返回“正确”结果的概率.

2.4.2 大矩阵乘法

本节考虑利用概率随机性来快速验证矩阵乘法的问题. 假设给定三个矩阵 $\mathbf{A}, \mathbf{B}, \mathbf{C} \in \{0, 1\}^{n \times n}$, 其中 n 非常大, 例如 $n \geq 10000000$, 我们研究的问题: 验证下面的矩阵乘法是否成立

$$\mathbf{AB} \stackrel{?}{=} \mathbf{C}.$$

若直接采用矩阵乘法计算 $\mathbf{AB}$, 然后再与矩阵 $\mathbf{C}$ 进行比较, 则计算复杂开销为 $O(n^3)$. 也可以采用更为精妙的算法, 比如采用分治策略, 目前最好的确定性算法的计算复杂开销为 $O(n^{2.37})$, 我们采用概率的随机方法进一步降低计算开销.

类似于验证多项式 $F(x) \equiv G(x)$ 的方法, 我们随机选取一个向量 $\bar{\mathbf{r}} = (r_1, r_2, \dots, r_n)^\top$, 其中元素 $r_1, r_2, \dots, r_n$ 都是从 $\{0, 1\}$ 中独立等可能随机选取所得. 下面验证

$$\mathbf{A}\bar{\mathbf{B}}\bar{\mathbf{r}} = \mathbf{A}(\mathbf{B}\bar{\mathbf{r}}) \stackrel{?}{=} \mathbf{C}\bar{\mathbf{r}}.$$

计算 $\mathbf{A}(\mathbf{B}\bar{\mathbf{r}})$ 和 $\mathbf{C}\bar{\mathbf{r}}$, 以及比较两个向量是否相等的计算复杂开销为 $O(n^2)$. 若 $\mathbf{A}(\mathbf{B}\bar{\mathbf{r}}) \neq \mathbf{C}\bar{\mathbf{r}}$ 则可以直接得到结果 $\mathbf{AB} \neq \mathbf{C}$; 而 $\mathbf{A}(\mathbf{B}\bar{\mathbf{r}}) = \mathbf{C}\bar{\mathbf{r}}$ 则不能直接得到结果 $\mathbf{AB} = \mathbf{C}$, 此时可以将上述过程独立地进行 k 次, 以此用较大的概率保证有 $\mathbf{AB} = \mathbf{C}$ 成立. 该过程被称为 Freivalds 算法, 如下所示:

输入: 矩阵 $\mathbf{A}, \mathbf{B}, \mathbf{C}$

输出: 是/否 %% 验证 $\mathbf{AB} \stackrel{?}{=} \mathbf{C}$

For $i = 1 : k$

 随机选择向量 $\bar{\mathbf{r}}_i = (r_{i1}, r_{i2}, \dots, r_{in})$, 其每个元素是从 $\{0, 1\}$ 独立等可能随机采样所得

 计算向量 $\bar{\mathbf{p}}_i = \mathbf{A}(\mathbf{B})\bar{\mathbf{r}}_i - \mathbf{C}\bar{\mathbf{r}}_i$

 If $\{\bar{\mathbf{p}}_i \text{ 不是零向量}\}$ then

 返回“否”

 EndIf

EndFor

返回“是”.

关于有效性, 若算法返回“否”, 则必有 $\mathbf{AB} \neq \mathbf{C}$, 因为找到了一个 $\bar{\mathbf{r}}$ 使得 $\mathbf{A}(\mathbf{B})\bar{\mathbf{r}} \neq \mathbf{C}\bar{\mathbf{r}}$ 成立; 若算法返回“是”, 则不一定有 $\mathbf{AB} = \mathbf{C}$ 成立, 但我们可以给出以较大的概率保证有 $\mathbf{AB} = \mathbf{C}$ 成立.

定理 2.5 设随机向量 $\bar{\mathbf{r}}_1, \bar{\mathbf{r}}_2, \dots, \bar{\mathbf{r}}_k \in \{0, 1\}^n$ 中每个元素都是从 $\{0, 1\}$ 独立等可能随机选取, 若 $\mathbf{AB} \neq \mathbf{C}$, 则有

$$P \left[\bigcap_{i=1}^k \{ \mathbf{A}(\mathbf{B})\bar{\mathbf{r}}_i = \mathbf{C}\bar{\mathbf{r}}_i \} \right] \leq \frac{1}{2^k}.$$

根据该定理可以选择 $k = \log_2 n$, 则 Freivalds 算法计算复杂度为 $O(n^2 \log n)$, 若算法返回“否”, 则有 $\mathbf{AB} \neq \mathbf{C}$; 若返回“是”, 则有 $P(\mathbf{AB} = \mathbf{C})$ 成立的概率超过 $1 - 1/n$.

证明 首先根据随机向量 $\bar{\mathbf{r}}_1, \bar{\mathbf{r}}_2, \dots, \bar{\mathbf{r}}_k$ 的独立同分布性有

$$P\left[\bigcap_{i=1}^k \{\mathbf{A}(\mathbf{B})\bar{\mathbf{r}}_i = \mathbf{C}\bar{\mathbf{r}}_i\}\right] = \prod_{i=1}^k P[\{\mathbf{A}(\mathbf{B})\bar{\mathbf{r}}_i = \mathbf{C}\bar{\mathbf{r}}_i\}] = (P[\{\mathbf{A}(\mathbf{B})\bar{\mathbf{r}}_1 = \mathbf{C}\bar{\mathbf{r}}_1\}])^k. \quad (2.1)$$

若 $\mathbf{AB} \neq \mathbf{C}$, 则必有 $\mathbf{D} = (d_{ij})_{n \times n} = \mathbf{AB} - \mathbf{C} \neq (0)_{n \times n}$, 此时不妨假设 $d_{11} \neq 0$. 随机向量 $\bar{\mathbf{r}}_1 = (r_{11}, r_{12}, \dots, r_{1n})^\top$ 中每个元素都是从 $\{0, 1\}$ 独立等可能随机选取, 由于结果返回“是”可知 $\mathbf{D}\bar{\mathbf{r}}_1 = 0$, 由此可得

$$d_{11}r_{11} + d_{12}r_{12} + \dots + d_{1n}r_{1n} = 0 \implies r_{11} = -\frac{d_{12}r_{12} + \dots + d_{1n}r_{1n}}{d_{11}}.$$

因此无论 $r_{12}, \dots, r_{1n}$ 取何值, 等式 $d_{11}r_{11} + d_{12}r_{12} + \dots + d_{1n}r_{1n} = 0$ 是否成立可根据 r_{11} 的值决定. 再根据 $P(r_{11} = 0) = P(r_{11} = 1) = 1/2$ 得到等式 $d_{11}r_{11} + d_{12}r_{12} + \dots + d_{1n}r_{1n} = 0$ 成立的概率不超过 $1/2$, 因此有

$$P[\{\mathbf{A}(\mathbf{B})\bar{\mathbf{r}}_1 = \mathbf{C}\bar{\mathbf{r}}_1\}] \leq 1/2.$$

结合 (2.1) 完成证明.

证明的思想又被称为 **延迟决策原理** (Principle of deferred decision), 当有多个随机变量解决一个问题时, 可以先着重考虑其中一个或一些变量, 而让其它剩余的变量保持随机性, 即延迟甚至不需考虑剩余变量对决策的影响. 在上面的证明过程中, 针对多个随机变量 $r_{11}, r_{12}, \dots, r_{1n}$, 我们着重考虑随机变量 r_{11} , 通过 r_{11} 概率的取值直接解决问题, 而没有考虑其它变量的可能性.

2.4.3 隐私问题的调查*

现实生活中的每个人都有一些隐私或秘密, 相关信息不希望被外人知晓, 然而对于一些具有社会普遍性的隐私问题, 我们需要对此进行一定的了解和调查, 例如在校大学生有抑郁倾向的同学占有多少比例, 家庭不和谐的同学占有多少比例, 等等. 这些信息属于个人隐私不便直接调查, 需要设计一种好的方案, 使被调查者愿意作出真实回答、又能较好地保护个人隐私.

经过多年研究与实践, 心理学家和统计学家设计了一种巧妙的方案, 核心是如下两个问题:

[问题 A:] 你的生日是否在 7 月 1 日之前?

[问题 B:] 你是否有抑郁的倾向?

再准备一个箱子, 里面装有 m 个白球和 n 个红球. 被调查者随机抽取一球, 若抽到白球回答问题 A, 否则回答问题 B. 在问卷的答案上只有两选项: “是”或“否”, 无论哪个问题都只需选择“是”或“否”, 最后将答卷放入一个投票箱内密封.

上述的抽球与回答过程都在一间无人的房间内进行, 任何外人都不知道被调查者抽到什么颜色的球, 也不知道被调查者的答案, 以此保护个人隐私. 如果向被调查者解释清楚了该调查方案并严格执行, 那么被调查者很容易确信他/她参加这次调查不会泄露个人隐私, 从而愿意配合调查.

当有 $N > 500$ 位学生参加调查后, 就可以打开投票箱进行统计. 设有 N_y 张答卷选择“是”, 根据频率与概率的关系有

$$P(\text{一个学生回答“是”}) \approx N_y/N.$$

设一个学生有抑郁倾向的概率为 p , 即

$$P(\text{一个学生回答“是”}|\text{红球}) = p.$$

不妨假设每个学生的生日是等可能事件, 因此一个学生在 7 月 1 日之前出生的概率为 $1/2$, 即

$$P(\text{一个学生回答“是”}|\text{白球}) = 1/2.$$

根据全概率公式有

$$P(\text{一个学生回答“是”}) = P(\text{一个学生回答“是”}|\text{红球})P(\text{红球}) + P(\text{一个学生回答“是”}|\text{白球})P(\text{白球}).$$

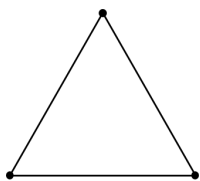
由此可得

$$\frac{N_y}{N} \approx \frac{m}{m+n} \times \frac{1}{2} + \frac{n}{m+n} \times p,$$

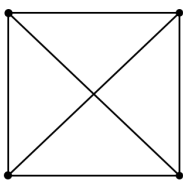
进一步估计出具有抑郁倾向的学生比例为 $p \approx (m+n)N_y/nN - m/2n$.

2.4.4 完全图着色*

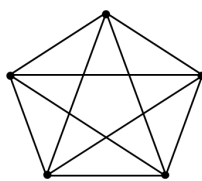
设平面上有 n 个顶点, 其中任意三个顶点不在同一条直线上, 用 $n(n-1)/2$ 条边将这些顶点连接起来的图称为 n 个顶点的 **完全图**, 例如三个、四个、五个顶点的完全图如下所示:



三个顶点的完全图



四个顶点的完全图



五三个顶点的完全图

将图中的每条边都分别染成红色或蓝色, 给定两正整数 $n \geq 10$ 和 $k > n/2$, 是否存在一种染色方法, 使得图上任意 k 个顶点相对应的 $k(k-1)/2$ 条边不是同一颜色?

我们利用概率的方法来求解该问题: 假设每条边等可能独立地被染成红色或蓝色, 即每条边为红色或为蓝色的概率均为 $1/2$. 从 n 个不同顶点中选出 k 个顶点有 $\binom{n}{k}$ 种不同的选法, 分别对应于 $\binom{n}{k}$ 个包含有 k 个顶点的子集, 这里将 k 个顶点的子集分别标号为 $1, 2, \dots, \binom{n}{k}$.

用 E_i 表示第 i 个子集中 $k(k-1)/2$ 条边染成相同颜色的事件, 根据题意可得

$$P(E_i) = 2(1/2)^{k(k-1)/2} \quad i = 1, 2, \dots, \binom{n}{k}.$$

若存在 k 个顶点, 其相应的 $k(k-1)/2$ 条边是同一颜色的事件可表示为 $\bigcup_{i=1}^{\binom{n}{k}} E_i$. 根据布尔不等式有

$$P\left(\bigcup_{i=1}^{\binom{n}{k}} E_i\right) \leq \sum_{i=1}^{\binom{n}{k}} P(E_i) = \binom{n}{k} (1/2)^{k(k-1)/2-1}.$$

当 $n \geq 10$ 和 $k > 2/n$ 时有 $P\left(\bigcup_{i=1}^{\binom{n}{k}} E_i\right) \leq 1$, 因此事件“完全图中任意 k 个顶点, 其相应的 $k(k-1)/2$ 条边不是同一颜色”的概率大于零. 这意味着至少存在一种染色方法, 使得对任意 k 顶点集合所对应的 $k(k-1)/2$ 边染色不全相同.

这种将概率用于求解纯粹确定性问题的方法称为 **概率化方法** (probabilistic method), 在计算机或人工智能中证明存在性时经常用到.

上述分析说明了完全图染色满足要求的存在性, 但并没有告诉我们如何涂颜色: 一种方法是随机涂色, 然后检查所涂的颜色是否满足所要求的性质; 若不成再重复进行直到成功为止.

习题

- 2.1 阐述独立与互不相容的关系.
- 2.2 若事件 A, B, C 独立, 证明: A 与事件 $B \cup C$ 独立.
- 2.3 设事件 A 和 B 满足 $P(A)P(B) > 0$, 证明: 若两事件独立则不互斥; 若两事件互斥则不独立.
- 2.4 小明同学参加时长为 1 小时的竞赛, 假设他在 $a \in [0, 1]$ 小时内完成竞赛的概率为 $a/2$. 已知小明在半小时后仍未完成, 求他最后要用完一小时的条件概率.
- 2.5 随机掷两次骰子, 已知第一次掷 6 点, 求两次点数之和不小于 10 的概率.
- 2.6 (三门问题) 在一电视节目中, 参赛者看到三扇关闭的门, 已知一门后面是汽车, 其它两门后面是山羊, 选中什么则获得什么, 主持人知道三门后有什么. 当参赛者选定一扇门但未开启, 此时节目主持人则开启剩下有山羊的一扇门. 问题: 若参赛者允许重新选择, 是否换一扇门?
- 2.7 已知事件 A 为病人被诊断为肝癌, 事件 C 为病人患有肝癌, $P(A|C) = 0.95$, $P(\bar{A}|\bar{C}) = 0.9$, $P(C) = 0.0004$. 求 $P(C|A)$.
- 2.8 书上的习题: 书 26 页到 28 页: 22, 27, 28, 30, 31, 32, 33, 37, 39.