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 发生的情况下事件 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

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

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

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

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 2.15 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 $(i = 0, 1, 2, \cdots, k)$. 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 $(i = 0, 1, 2, \cdots, k)$. 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 $k \to +\infty$ 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 $A, B, C \in \Sigma$ 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 $A_1, A_2, \cdots, A_n \in \Sigma$ 中任意 k 个事件是相互独立的 $(k \geqslant 2)$, 即满足

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

则称事件 $A_1, A_2, \cdots, A_n$ 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

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

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

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

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

$$P(\bar{A}_1 \cup \bar{A}_2 \cup \cdots \cup \bar{A}_n) = 1 - P(A_1A_2\cdots A_n) = 1 - p_1p_2\cdots p_n.$$

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 $A_1, A_2, \cdots, A_n, \cdots$ 独立且每事件发生的概率 $P(A_i) = p > 0$ 非常小, 则有

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

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

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

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 $p_1 + p_2 + \cdots + p_m = 1$. 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

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

同理可得

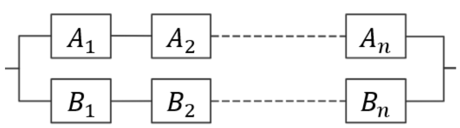
$$P(A_i\cup A_j)=1-P(\overline{A_i}\cap\overline{A_j})=1-(1-p_i-p_j)^n\ .$$

利用容斥原理和条件概率的定义有

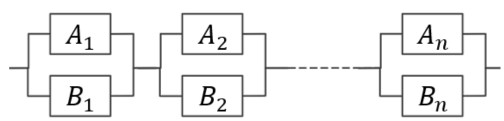
$$\begin{aligned}P(A_i\cup A_j) &= \frac{P(A_iA_j)}{P(A_j)}=\frac{P(A_i)+P(A_j)-P(A_i\cup A_j)}{P(A_j)}\\ &= \frac{1-(1-p_i)^n-(1-p_j)^n+(1-p_i-p_j)^n}{1-(1-p_j)^n}\ .\end{aligned}$$

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.



系统 I



系统 II

解 用事件 A_i 和 B_i 表示图中所对应的元件正常工作 $(i=1,2,,,n)$. 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

$$\begin{aligned}&P((A_1A_2\cdots A_n)\cup(B_1B_2\cdots B_n))\\ &= P(A_1A_2\cdots A_n)+P(B_1B_2\cdots B_n)-P(A_1A_2\cdots A_nB_1B_2\cdots B_n)=2p^n-p^{2n}=p^n(2-p^n).\end{aligned}$$

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

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

利用数学归纳法可证明当 $n\geqslant 2$ 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 $\ln(1-x) \leqslant -x$, 于是得到

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

分别对上式两边取极限 $n \to +\infty$ 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 $\lim_{n\to\infty} \sum_{i=n}^{\infty} P(A_i) = 0$. 根据题意可知 $A \subseteq \bigcup_{i=n}^{+\infty} A_i$, 利用Union bounds有

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

上式两边同时对 $n \to +\infty$ 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 $m > n \geqslant 1$, 利用德摩根律和独立性假设有

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

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

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

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 $[100d]=\{1,2,\cdots,100d\}$ 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
$$P[F(r)=G(r)]\leqslant d/100d=1/100.$$

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 $r_1,r_2,\cdots,r_k$. 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 $F(r_1)=G(r_1),F(r_2)=G(r_2),\cdots,F(r_k)=G(r_k)$ 的概率, 根据事件的独立性与前面的分析有

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

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 $n \geqslant 10000000$, 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 $\bar{\mathbf{r}} = (r_1, r_2, \cdots, r_n)^\top$, 其中元素 $r_1, r_2, \cdots, r_n$ 都是从 $\{0,1\}$ 中独立等可能随机选取所得. 下面验证

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

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 $\bar{\mathbf{r}}_i = (r_{i1}, r_{i2}, \cdots, r_{in})$, 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 $\{\bar{\mathbf{p}}_i$ 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 $\bar{\mathbf{r}}_1, \bar{\mathbf{r}}_2, \cdots, \bar{\mathbf{r}}_k \in \{0,1\}^n$ 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

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

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 $\bar{\mathbf{r}}_1,\bar{\mathbf{r}}_2,\cdots,\bar{\mathbf{r}}_k$ 的独立同分布性有

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

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 $\bar{\mathbf{r}}_1=(r_{11},r_{12},\cdots,r_{1n})^\top$ 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

$$d_{11}r_{11}+d_{12}r_{12}+\cdots+d_{1n}r_{1n}=0 \quad\Longrightarrow\quad r_1=-\frac{d_{12}r_{12}+\cdots+d_{1n}r_{1n}}{d_{11}}\ .$$

因此无论 $r_{12},\ldots,r_{1n}$ 取何值, 等式 $d_{11}r_{11}+d_{12}r_{12}+\cdots+d_{1n}r_{1n}=0$ 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 $d_{11}r_{11}+d_{12}r_{12}+\cdots+d_{1n}r_{1n}=0$ 成立的概率不超过 $1/2$, 因此有

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

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 $r_{11},r_{12},\cdots,r_{1n}$, 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

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

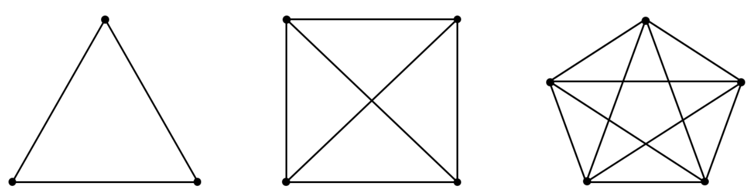
由此可得

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

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 $n \geqslant 10$ 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 $1, 2, \cdots, \binom{n}{k}$.

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

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

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

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

当$n \geqslant 10$ 和 $k > 2/n$ 时有 $P\left(\bigcup_{i=1}^{\binom{n}{k}} E_i\right) \leqslant 1$, 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.