人工智能中的数学讲义

方聪

北京大学

摘要

本讲义收录了人工智能中的数学课程中的主要概念与课程习题。概率与统计讲义内容摘录于陈家鼎、郑忠国《概率与统计》教材与复熹和张原概率与统计课程课件。图论内容摘录于耿素云、屈婉玲、王捍贫《离散数学教程》。本讲义版权归上述作者，不会出版。讲义仅供于上该课程的同学们学习参考，讲义的错误会不断修正。感谢张乙沐、张海涵等助教对讲义整理的帮助。

目　录

第一章　随机事件与概率

- 随机事件、样本空间、样本点概念（掌握）
- 事件与概率的运算（掌握）
- 概率公理定义（了解）
- 古典概型（掌握）
- 条件概率、独立性（掌握）
- 贝叶斯公式与全概率公式（掌握）

1.1 随机事件及其运算

1.1.1 随机事件

在考虑一个未来事件是否会发生的时候, 人们常常关心该事件发生的可能性大小, 而概率正是用来刻画这种可能性的度量. 为了引入概率概念, 需要先明确研究的基本对象——随机实验及其可能结果.

定义 1.1.1 (随机实验)**.** 凡满足以下条件的过程称为**随机实验**:

(1) 在相同条件下可以重复进行;

(2) 可能出现两个或两个以上不同的结果;

(3) 在实验进行之前不能确定哪一个结果会出现.

随机实验通常简称为**实验**, 可记作 E.

虽然单次实验的结果是不能提前确定的, 但所有可能的结果可以预先知道. 为了研究随机现象, 我们不仅需要考察这些结果的总体, 还要对每一个结果本身加以区分, 这就引出了样本空间和样本点的概念.

定义 1.1.2 (样本空间和样本点)**.** 随机实验 E 中所有可能结果组成的集合称为实验 E 的**样本空间**, 记为 Ω. 样本空间中的元素称为**样本点**, 记为 ω.

为了便于理解这一概念, 我们来看几个简单的例子:

- E_1: 抛掷硬币, 观察正面 H, 反面 T 出现的情况.

$$\Omega_1 = \{H, T\}.$$

- E_2: 抛掷一枚硬币 3 次, 观察正面 H, 反面 T 出现的情况.

$$\Omega_2 = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}.$$

- E_3: 抛掷一枚硬币 3 次, 观察正面出现的次数.

$$\Omega_3 = \{0, 1, 2, 3\}.$$

在此基础上, 我们往往关心其中某些结果的集合, 这就引出了随机事件的概念.

定义 1.1.3 (随机事件). 随机实验的若干样本点组成的集合称为**随机事件**, 简称为**事件**, 常用 $A, B, C, \cdots$ 表示.

例如, 设实验 E 为抛掷一枚骰子, 样本空间 $\Omega = \{1, 2, 3, 4, 5, 6\}$, 则事件 $A =$“出现奇数点” (即 $A = \{1, 3, 5\}$) 是样本空间 $\Omega = \{1, 2, 3, 4, 5, 6\}$ 的一个子集.

定义 1.1.4 (不可能事件与必然事件). 样本点组成的集合为空集 $\varnothing$ 时, 称事件为**不可能事件**; 样本点组成的集合为样本空间全集 Ω 时, 称事件为**必然事件**, 不可能事件与必然事件是随机事件的两种特殊情形.

定义了事件之后, 我们还需要定量描述事件发生的可能性, 因此我们需要定义事件的概率, 建立在随机实验上的概率定义如下:

定义 1.1.5 (频率与概率). 设 μ 是 n 次实验中事件 A 发生的次数, 则事件 A 发生的**频率**为 $\frac{\mu}{n}$. 若随着实验次数 n 增大, 频率 $\frac{\mu}{n}$ 最终稳定地在某一数值 p 附近摆动, 则该 p 值称为该事件的**概率**, 记为 $P(A) = p$.

例如, 进行实验“投一枚硬币” n 次, 出现正面朝上的次数为 μ_n, 则有 $\lim\limits_{n\to\infty} \frac{\mu_n}{n} = p$, 其中 p 就是实验“投一枚硬币”中, 事件“正面朝上”的概率.

根据定义, 频率 $\frac{\mu}{n}$ 总介于 $0, 1$ 之间, 因此概率 $P(A) = p$ 也存在基本性质:

$$0 \leqslant P(A) \leqslant 1.$$

现实世界中, 我们也会遇到无法进行随机实验的情况, 如医生告诉病人手术成功的“概率”, 此类事件的“概率”是人们根据已有的知识和经验对该事件发生的可能性所给出的个人信念. 这种方式定义的概率被称为“主观概率”, 表示对人们事件发生的置信度.

不论是由频率定义的概率, 还是主观概率, 都是对事件发生可能性大小的度量. 大概率事情易发生, 小概率事情不易发生.

1.1.2 事件的运算

在随机实验中存在多个事件, 其中有些较为简单, 有些则较为复杂. 分析事件之间的关系, 特别是揭示简单事件与复杂事件之间的联系, 对于确定某些事件的概率具有重要意义. 因此, 需要讨论事件的关系与运算, 此处的“运算”是指从一些已知事件出发构造出新事件的规则.

1.1.2.1　事件的包含与相等

先从事件之间最基本的关系出发, 我们定义事件的包含和相等关系.

定义 1.1.6 (事件的包含)**.** 设有事件 A 和事件 B, 如果 A 发生, 则 B 必发生, 那么称事件 B **包含**事件 A(或称事件 A 在 B 中), 并记为 $A \subset B$ (或 $B \supset A$).

定义 1.1.7 (事件的相等)**.** 如果事件 A 包含事件 B, 同时事件 B 包含事件 A, 则事件 A 和事件 B **相等**, 并记为 $A = B$.

1.1.2.2　事件的并和交

接下来考虑最基本的事件运算, 我们定义事件的并和交运算.

定义 1.1.8 (事件的并)**.** 设 A 和 B 都是事件, 则 "A 或 B" 表示这样的事件 C: C 发生当且仅当 A 或 B 中至少有一个发生, 该事件 C 叫做 A 与 B 的**并**, 记为 $A \cup B$.

例 1.1.1 (对应郑书例 2.1)**.** 在桌面上, 投掷两枚匀称的硬币, A 表示 "恰好一枚国旗朝上" , B 表示 "两枚国旗朝上" , C 表示 "至少一枚国旗朝上" , 则 $C = A \cup B$.

以下恒记必然事件为 U, 不可能事件为 V.

对于并运算, 有以下性质:

$$A \cup B = B \cup A,$$
$$A \cup U = U,\ A \cup V = A.$$

定义 1.1.9 (事件的交)**.** 设 A 和 B 都是事件, 则 "A 且 B" 表示这样的事件 C: C 发生当且仅当 A 和 B 都发生, 该事件 C 叫做 A 与 B 的**交**, 记为 $A \cap B$, 也简记为 AB.

在例 1.1.1 中, $A \cap C = A$, $B \cap C = B$, $A \cap B = V$.

对于交运算, 有以下性质:

$$A \cap B = B \cap A,$$
$$A \cap U = A,\ A \cap V = V.$$

1.1.2.3 事件的余和差

接下来, 我们定义事件的余和差运算. 与此同时, 可以得到事件的另一种基本关系——对立.

定义 1.1.10 (事件的余)**.** 设 A 是事件, 称 “非 A” 是 A 的**对立事件** (或称**余事件**), 其含义为: “非 A” 发生当且仅当 A 不发生, 常常用 $\overline{A}$ 表示 “非 A” , 也用 A^c 表示 “非 A” .

由定义知 $\overline{(\overline{A})} = A, \overline{U} = V, \overline{V} = U$.

定义 1.1.11 (事件的差)**.** 设 A 和 B 都是事件, 则两个事件的**差** “A 减去 B” 表示这样的事件 C: C 发生当且仅当 A 发生而 B 不发生, 该事件 C 记为 $A - B$(或 $A \setminus B$).

由定义知, $A - B = A \cap \bar{B}$.

事件的关系可用画图法直观理解:

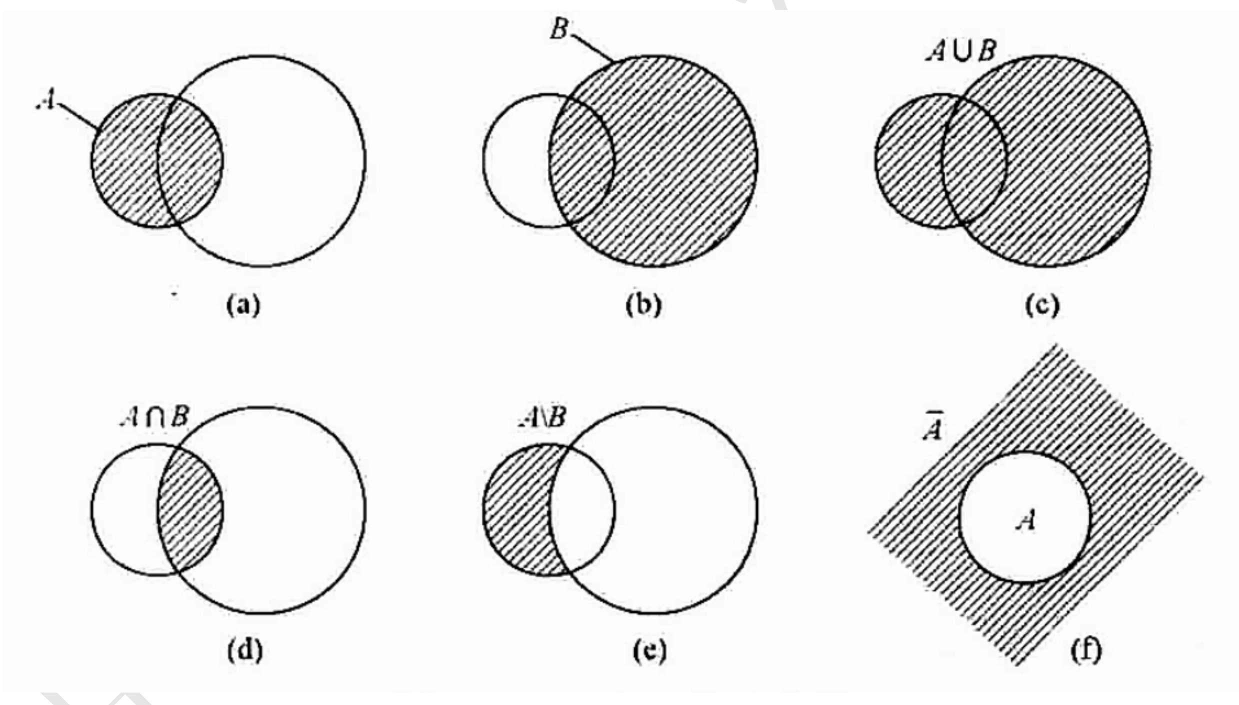

图 *1.1:* 事件运算示意图

1.1.2.4 事件运算的性质

事件的基本运算还有以下性质:

- $A \cup (B \cup C) = (A \cup B) \cup C$ ("并"的结合律);
- $A \cap (B \cap C) = (A \cap B) \cap C$ ("交"的结合律);
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ (分配律);
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ (分配律);
- $A \cup A = A$, $A \cap A = A$;
- $\overline{A \cup B} = \overline{A} \cap \overline{B}$ (对偶律);
- $\overline{A \cap B} = \overline{A} \cup \overline{B}$ (对偶律).

由于并和交有结合律, 我们可以定义多个事件的并和交:

定义 1.1.12 (多个事件的并和交)**.** 设 $A_1, A_2, \cdots, A_n$ 是 n 个事件. "$A_1, A_2, \cdots, A_n$"的并是指这样的事件: 它发生当且仅当 $A_1, A_2, \cdots, A_n$ 中至少一个发生, 常用 $\bigcup\limits_{i=1}^{n} A_i$ 表示; "$A_1, A_2, \cdots, A_n$"的交是指这样的事件: 它发生当且仅当 $A_1, A_2, \cdots, A_n$ 都发生, 常用 $\bigcap\limits_{i=1}^{n} A_i$ 或 $A_1 A_2 \cdots A_n$ 表示.

进一步, 对于可列无穷个事件, 也可以这样定义并和交:

定义 1.1.13. 设 $A_1, A_2, \cdots, A_i, \cdots$ 是一列事件. 诸 A_i 的并 B 是指这样的事件: B 发生当且仅当这些 $A_i (i = 1, 2, \cdots)$ 中至少一个发生, 记为 $B = \bigcup\limits_{i=1}^{\infty} A_i$ 或 $B = A_1 \cup A_2 \cup \cdots$; 诸 A_i 的交 C 是指这样的事件: C 发生当且仅当这些 $A_i (i = 1, 2, \cdots)$ 都发生, 记为 $C = \bigcap\limits_{i=1}^{\infty} A_i$ 或 $C = A_1 A_2 \cdots$.

另外, 若事件集不可列, 则有更一般的定义方式:

定义 1.1.14. 设 $\{A_a, a \in \Gamma\}$ 是一族事件 (其中 Γ 是任何非空集, 每个 $a \in \Gamma$ 对应一个事件 A_a). 事件 A_a 的"并"是指这样的事件 B: B 发生当且仅当至少一个 A_a 发生, 常记为 $B = \bigcup\limits_{a \in \Gamma} A_a$; 事件 A_a 的"交"是指这样的事件 C: C 发生当且仅当 A_a 都发生, 常记为 $C = \bigcap\limits_{a \in \Gamma} A_a$.

例 1.1.2. 取 $X \in \mathbb{R}$, 事件 A_i 为 $X \in [\frac{1}{i+1}, \frac{1}{i}]$, 事件 B_i 为 $X \in [0, \frac{1}{i}]$. 则事件 $\overset{n}{\underset{i=1}{\cup}} A_i$ 发生等价于 $X \in [\frac{1}{n+1}, 1]$, 事件 $\overset{n}{\underset{i=1}{\cap}} B_i$ 发生等价于 $X \in [0, \frac{1}{n}]$. 进而当 $n \to \infty$ 时事件 $\overset{\infty}{\underset{i=1}{\cup}} A_i$ 发生等价于 $X \in (0, 1]$, 事件 $\overset{\infty}{\underset{i=1}{\cap}} B_i$ 发生等价于 $X = 0$.

不难验证, 对可列个事件的并和交有以下规律:

- $A \cup (\bigcap\limits_{i=1}^{\infty} B_i) = \bigcap\limits_{i=1}^{\infty} (A \cup B_i)$ (分配律);
- $A \cap (\bigcup\limits_{i=1}^{\infty} B_i) = \bigcup\limits_{i=1}^{\infty} (A \cap B_i)$ (分配律);
- $\overline{(\bigcup\limits_{i=1}^{\infty} A_i)} = \bigcap\limits_{i=1}^{\infty} \overline{A_i}$ (对偶律);
- $\overline{(\bigcap\limits_{i=1}^{\infty} A_i)} = \bigcup\limits_{i=1}^{\infty} \overline{A_i}$ (对偶律).

1.1.2.5 互不相容的事件

定义了事件的交后, 我们希望定义事件的交为空的关系, 称这种关系为互不相容 (互斥).

定义 1.1.15 (互不相容的事件)**.** 如果事件 A 和事件 B 不能都发生, 即 $A \cap B = V$, 则称 A 和 B 是**互不相容**的事件 (也称**互斥**的事件). 对多个事件 $A_1, \cdots A_n$, 若对任何 $i \neq j (i, j = 1, \cdots n)$ 有 A_i 与 A_j 互不相容, 则称事件 $A_1, \cdots A_n$ 互不相容.

例如, 抛掷两枚硬币, 事件"恰好一枚国徽朝上"和事件"两枚都是国徽朝上"是互不相容的. 不难看出, 对任何事件 A, A 和 $\overline{A}$ 是互不相容的.

1.1.2.6 概率的加法公式

经过上述讨论, 不难发现, 事件的关系和运算本质上就是集合的关系和运算.

定义了一些基本的关系和运算后, 我们回到最初的目的, 通过简单事件的概率计算复杂事件的概率. 因此, 我们给出概率的加法公式.

对于互不相容的事件, 从实际经验出发, 我们有以下基本事实:

公理 1.1.1 (概率的加法公式)**.** 若 $A_1, A_2, \ldots$ 互不相容, 则:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i);$$

由此, 我们可以证明如下性质:

推论 1.1.1. 对于任意两个事件 A 和 B, 有 $P(A \cup B) = P(A) + P(B) - P(AB)$.

1.2 概率的公理化定义（了解）

前面我们简单讨论了事件以及概率的概念, 本节中我们按照公理化方法给概率下一个严格定义.

定义 1.2.1 (概率空间子类)**.** 设 Ω 为样本空间, $\mathscr{F}$ 为 Ω 的一些子集构成的集类. 若 $\mathscr{F}$ 满足以下三个条件:

(1) $\Omega \in \mathscr{F}$;

(2) $A \in \mathscr{F} \Rightarrow A^c \in \mathscr{F}$;

(3) $\{A_n\}_{n\in\mathbb{N}} \subsetneqq \mathscr{F} \Rightarrow \bigcup\limits_{n=1}^{\infty} A_n \in \mathscr{F}$,

则称 $\mathscr{F}$ **为概率空间子类**.

以下是几个概率空间子类的例子:

- $\mathscr{F} = \{\varnothing, \Omega\}$ 平凡概率空间子类;
- $\mathscr{F} = \{\varnothing, \Omega, A, \overline{A}\}$ 包含 A 的最小概率空间子类;
- $\mathscr{F} = \{A | A \subset \Omega\}$ Ω 上的最大概率空间子类;
- $\Omega = \{\omega_1, \cdots, \omega_n\}$, 则 Ω 所有子集构成的概率空间子类共有 2^n 个元素.

定义 1.2.2 (概率空间)**.** 设 $\mathscr{F}$ 是满足上述条件的概率空间子集类. 概率 $P = P(\cdot)$ 是 $\mathscr{F}$ 上面定义的实值函数, 满足:

(1) 非负性: $P(A) \geqslant 0$ 对于一切 $A \in \mathscr{F}$;

(2) 规范性: $P(\Omega) = 1$;

(3) 可列可加性: 若 $A_n \in \mathscr{F}(n = 1, 2, \cdots)$ 两两不相交, 则

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

满足以上条件的 $(\Omega, \mathscr{F}, P)$ 称为**概率空间**.

例如假定 $\Omega = \{\omega_1, \cdots, \omega_n\}$, $\mathscr{F}$ 为全体子集构成的概率空间子类. 设 $p_1, \cdots, p_n$ 为 n 个非负实

数, 且满足 $\sum\limits_{i=1}^{n} p_i = 1$. 令

$$\mathbb{P}(\varnothing) = 0, \quad \mathbb{P}(A) = \sum_{j=1}^{k} p_{i_j}, \quad A = \{\omega_{i_1}, \cdots, \omega_{i_k}\}, k = 1, \cdots, n,$$

则 $\mathbb{P}$ 为 $(\Omega, \mathscr{F})$ 上概率.

概率 P 有以下性质:

(1) $P(\varnothing) = 0$;

(2) 若 $A \in \mathscr{F}$, 则 $P(A^c) = 1 - P(A)$;

(3) 若 $A_1, \cdots, A_n$ 都属于 $\mathscr{F}$ 且两两不相交, 则

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i); \tag{1.2.1}$$

(4) 若 $A \subset B$, $A \in \mathscr{F}$, $B \in \mathscr{F}$, 则 $P(A) \leqslant P(B)$, 且

$$P(B - A) = P(B) - P(A); \tag{1.2.2}$$

(5) 若 $A_n \subset A_{n+1}$, $A_n \in \mathscr{F}(n = 1, 2, \cdots)$, 则

$$P(\bigcup_{n=1}^{\infty} A_n) = \lim_{n \to \infty} P(A_n); \tag{1.2.3}$$

(6) 若 $A_n \supset A_{n+1}$, $A_n \in \mathscr{F}(n = 1, 2, \cdots)$, 则

$$P(\bigcap_{n=1}^{\infty} A_n) = \lim_{n \to \infty} P(A_n); \tag{1.2.4}$$

(7) 若 $A_n \in \mathscr{F}(n = 1, 2, \cdots)$, 则

$$P(\bigcup_{n=1}^{\infty} A_n) \leqslant \sum_{n=1}^{\infty} P(A_n). \tag{1.2.5}$$

1.3 古典概型

在概率的频率定义中, 为了确定一个事件 A 的概率, 需要在条件 S 下大量重复实验. 然而, 在某些情形下, 只需要利用人们的共识或者经验就可以无异议地确定事件的概率. 例如, 在抛硬币的例子中, 即使不进行重复实验, 人们也会认可"国徽向上"和"国徽向下"的概率都是 1/2. 这是因为问题本身存在对称性 (硬币均匀) . 再看另外一个例子, 在 100 件产品中有不合格品 5 件, 从中任取 (这里隐含了每个产品被抽到的机会相同) 一件, 则抽到不合格品的概率是 1/20.

上面例子的一个共同特点, 是可以通过"对称性"确定事件 $A_1, A_2, \ldots, A_n$ 是等可能的, 并且这些事件没有重复和遗漏. 我们将这样的事件称为**基本事件**, 并把这样的概率模型称为古典概型.

定义 1.3.1. 若随机现象有如下两个特征:
(1) 在实验中它的全部可能性只有有限个;
(2) 基本事件发生或出现是等可能的;
则称其对应的数学模型为**古典概型**.

也可以用公理化的方法定义古典概型. 取

$$\Omega = \{\omega_1, \omega_2, \cdots, \omega_n\}, \quad \mathscr{F} = \{A | A \subset \Omega\},$$

令 P 为 $(\Omega, \mathscr{F})$ 上的概率测度, 满足

$$P(\{\omega_1\}) = \cdots = P(\{\omega_n\}).$$

则 $(\Omega, \mathscr{F}, P)$ 为古典概型对应的概率空间.

根据古典概型的定义, 容易得到**古典概型的概率计算公式**. 设事件 $A = \{\omega_{i_1}, \cdots, \omega_{i_k}\} \in \mathscr{F}$, 利用概率的有限可加性可知:

$$P(A) = \sum_{j=1}^{k} P(\{\omega_{i_j}\}) = \frac{k}{n} = \frac{|A|}{|\Omega|}. \tag{1.3.1}$$

在利用式 (1.3.1) 计算概率时, 主要的困难是确定事件中包含基本事件的数量, 因此需要利用排列组合的公式. 排列和组合给出了两种经典的计数问题的结果, 更复杂的计数问题往往可以转化为这两类问题, 这种技巧需要同学们通过练习掌握.

排列: 从含有 n 个不同元素的总体中, 无放回地抽取 r 个进行排列, 其总数为

$$A_n^r := n(n-1)\cdots(n-r+1).$$

当 $r = n$ 时, 称为全排列, 此时共有 $A_n^n = n!$ 种排列方式. 若考虑有放回的抽取, 则共有 n^r 种排列方式.

组合: 从 n 个不同元素中取出 r 个而不考虑其顺序, 称为组合, 其总数为

$$C_n^r := \frac{n!}{r!(n-r)!} = \frac{A_n^r}{r!}.$$

下面是一些其他常见的组合问题的结果:

(1) 把 n 个不同元素分成 k 个部分, 且第 i 个部分有 r 个元素, $1 \leqslant i \leqslant k$, 且 $r_1+r_2+\cdots+r_k=n$, 则有 $\frac{n!}{r_1!r_2!\cdots r_k!}$ 种方法;

(2) 把 n 个元素全部带有标注, 其中 n_1 个带标注 1, n_2 个带标注 2, $\cdots$, n_k 个带标注 k. 现在从此 n 个元素中取出 r 个, 使得带有标注 i 的元素有 r_i 个, 其中 $1 \leqslant i \leqslant k$ 且 $r_1+r_2+\cdots+r_k=r$. 则不同取法的总数为 $C_{n_1}^{r_1}C_{n_2}^{r_2}\cdots C_{n_k}^{r_k}$.

(3) 从 n 个不同元素中有重复的取出 r 个, 不计顺序, 则不同的取法有 C_{n+r-1}^r (有重复组合数)

组合数公式: 对一切正整数 a,b,

$$\sum_{i=0}^{n} C_a^i C_b^{n-i} = C_{a+b}^n,$$

约定当 $k>n$ 时, $C_n^k=0$. 特别地,

$$\sum_{i=0}^{n} (C_n^i)^2 = C_{2n}^n.$$

例 1.3.1 (对应郑书例 3.1)**.** 某人同时抛掷两枚骰子, 问: 得到 7 点 (两颗骰子的点数之和) 的概率是多少?

解. 我们用甲乙分别表示这两颗骰子, 每颗骰子共有 6 种可能的点数: 1, 2, 3, 4, 5, 6, 两颗骰子共有 6×6=36 种可能结果: $(i,j)(i=1,\cdots,6)(j=1,\cdots,6)$, 这里 i 表示骰子甲的点数, j 表示骰子乙的点数, 显然这些结果出现的机会是相等的, 它们构成了等概完备事件组, 事件 "得到 7 点" 由 6 种结果 (基本事件) 组成: $(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)$, 故事件 "得到 7 点" 的概率为 $\frac{6}{36}=\frac{1}{6}$ □

例 1.3.2. 甲口袋有 5 个白球, 3 个黑球, 乙口袋中有 4 个白球, 6 个黑球, 从两个口袋中各任取一球, 求取到的两个球颜色相同的概率.

解. 从两个口袋中各取一球, 共有 $C_8^1C_{10}^1$ 种等可能取法. 两球颜色相同可能情况为: 从甲乙口袋均取出白球, 从甲乙口袋均取出黑球, 共有 $C_5^1C_4^1+C_3^1C_6^1$ 种取法, 于是

$$P(\text{取到的两个球颜色相同}) = \frac{C_5^1C_4^1+C_3^1C_6^1}{C_8^1C_{10}^1} = \frac{19}{40}.$$

□

例 1.3.3 (巴拿赫问题)**.** 某数学家有两盒火柴, 每盒有 n 根, 每次使用时, 他任取一盒并从中抽出一根, 问他发现一盒空而同时另一盒还有 $r(0 \leqslant r \leqslant n)$ 的概率为多少 (发现为空表示最后一次抽到空盒)?

解. 设两盒火柴分别为 A, B, 由对称性, 所求概率为事件 $E=$ "发现 A 盒空而 B 盒还有 r 根" 的概率的 2 倍. 先计算样本空间中的样本点个数, 由于共取了 $2n-r+1$ 次, 故有 2^{2n-r+1} 个样本点. 考察事件 E, 等效为前 $2n-r$ 次 A 盒恰好取 n 次, 次序不论, 最后一次必定取到 A 盒, 此种样本点共有 C_{2n-r}^{n} 个, 因此

$$P(E)=\frac{C_{2n-r}^{n}}{2^{2n-r+1}}.$$

所求概率为 $\frac{C_{2n-r}^{n}}{2^{2n-r}}$. □

1.4 条件概率与独立性

1.4.1 条件概率

到现在讨论的概率都是相对于某一确定的条件 S, 我们同样可以在 S 之外附加新的条件. 更确切的说, 设 A 和 B 都是条件 S 下的事件, 我们希望研究在 B 发生的条件下, 事件 A 发生的概率, 记作 $P(A|B)$. 和定义 1.1.5 一样, 不断进行实验并将频率 μ_B/n_B 的 "极限" 定义为概率. 不过不同的是这里的 μ_B 代表 A 和 B 同时发生的次数, 而 n_B 是条件 B 发生的次数 (即忽略条件 B 没有发生的情况) . 在分子分母上同时除以总实验次数 n, 得到条件概率

$$P(A|B)\approx\frac{\mu_B/n}{n_B/n}\approx\frac{P(AB)}{P(B)},$$

这里 $\approx$ 表示当实验次数足够大时在概率值的附近摆动. 为了数学上的严谨性, 我们采用如下的定义.

定义 1.4.1. 设 $(\Omega,\mathscr{F},P)$ 为概率空间, 事件 $A,B\in\mathscr{F}$ 满足 $P(B)>0$. 称

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

为 B 发生条件下 A 发生的**条件概率**.

条件概率 $P(\cdot|B)$ 为 $(\Omega,\mathscr{F})$ 上的概率, 即满足定义 1.2.2:

(1) 对任意事件 $A\in\mathscr{F}$, 有 $P(A|B)\geqslant 0$;

(2) $P(\Omega|B)=1$;

(3) 对任意两两不相交的事件序列 $\{A_n\}_{n=1}^{\infty} \subset \mathscr{F}$, 有

$$P\left(\bigcup_{n=1}^{\infty} A_n \middle| B\right) = \sum_{n=1}^{\infty} P(A_n|B).$$

由条件概率的定义, 容易得到, $P(B|\Omega) = P(B)$. 将条件概率的定义改写可以得到**乘法公式:**

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

对于 n 个事件, 可以归纳得到**乘法公式的推广:**

$$P(A_1A_2\cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1}),$$

其中 $P(A_1A_2\cdots A_{n-1}) > 0$. 表明上看, 乘法公式只是将 $P(AB)$ 的问题转化成了计算 $P(B|A)$, 而依定义计算 $P(B|A)$ 又需要 $P(AB)$ 的值, 因此并没有解决问题. 但实际上, $P(B|A)$ 的值有时可以通过条件概率的直观意义得到, 从而求出复杂事件 AB 的概率, 这种应用可以参考下面的例题.

例 1.4.1. 将 52 张扑克牌 (不含大王、小王) 随机地分为 4 堆, 每堆 13 张, 问: 各堆都含有 A 牌 (即 1 点) 的概率是多少?

解. 将 4 堆扑克牌编号: 第 1 堆, 第 2 堆, 第 3 堆, 第 4 堆, 用 A_1, A_2, A_3, A_4 依次表示 4 个 A 牌, 设 i_1, i_2, i_3, i_4 是 1, 2, 3, 4 的一个排列, 令 $E_{i_1i_2i_3i_4}$ = “第 i_1 堆有 A_1 但没有 A_2, A_3, A_4, 第 i_2 堆有 A_2 但没有 A_1, A_3, A_4, 第 i_3 堆有 A_3 但没有 A_1, A_2, A_4, 第 i_4 堆有 A_4 但没有 A_1, A_2, A_3”, E = “各堆都含有 A”, 则

$$E = \bigcup_{i_1i_2i_3i_4} E_{i_1i_2i_3i_4},$$

这些事件两两不相容, 易知 $P(E) = 4!P(E_{1234})$, 令 $E_k = \{$ 第 k 堆含有 A_k 但不含有其他的 $A_j (j \neq k)\} \quad (k = 1, 2, 3, 4)$, 则

$$P(E_{1234}) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2)P(E_4|E_1E_2E_3).$$

易知

$$P(E_1) = C_{48}^{12}/C_{52}^{13}, \quad P(E_2|E_1) = C_{36}^{12}/C_{39}^{13},$$

$$P(E_3|E_1E_2) = C_{24}^{12}/C_{26}^{13}, \quad P(E_4|E_1E_2E_3) = 1,$$

于是

$$P(E_{1234}) = \frac{C_{48}^{12}C_{36}^{12}C_{24}^{12}}{C_{52}^{13}C_{39}^{13}C_{26}^{13}} = \frac{13^4}{52\times51\times50\times49},$$

$$P(E) = 4!P(E_{1234}) \approx 0.105.$$

□

例 1.4.2 (罐子模型)**.** 设罐中有 b 个黑球, r 个红球, 每次随机取出一个球, 取出后将原球放回, 还加进 c 个同色球和 d 个异色球, 记 B_i 为 "第 i 次取出的是黑球", R_j 为 "第 j 次取出的是红球". 若连续从罐中取出三个球, 其中有两个红球, 一个黑球, 则由乘法公式我们可得

$$P(B_1R_2R_3) = P(B_1)P(R_2|B_1)P(R_3|B_1R_2) = \frac{b}{b+r}\cdot\frac{r+d}{b+r+c+d}\cdot\frac{r+d+c}{b+r+2c+2d},$$

$$P(R_1B_2R_3) = P(R_1)P(B_2|R_1)P(R_3|R_1B_2) = \frac{r}{b+r}\cdot\frac{b+d}{b+r+c+d}\cdot\frac{r+d+c}{b+r+2c+2d},$$

$$P(R_1R_2B_3) = P(R_1)P(R_2|R_1)P(B_3|R_1R_2) = \frac{r}{b+r}\cdot\frac{r+c}{b+r+c+d}\cdot\frac{b+2d}{b+r+2c+2d},$$

以上概率与黑球在第几次被抽出有关. 罐子模型也称波利亚 (Polya) 模型, 这个模型的各种变化如下:

(1) 当 $c=-1$, $d=0$ 时, 为不返回抽样, 此时前次抽取结果会影响后次抽取结果, 但只要抽取的黑球和红球个数确定, 则概率不依赖其抽出球的次序, 有

$$P(B_1R_2R_3) = P(R_1B_2R_3) = P(R_1R_2B_3) = \frac{br(r-1)}{(b+r)(b+r-1)(b+r-2)}.$$

(2) 当 $c=0$, $d=0$ 时, 为返回抽样, 此时前次抽取结果不会影响后次抽取结果, 上述三种概率相等, 有

$$P(B_1R_2R_3) = P(R_1B_2R_3) = P(R_1R_2B_3) = \frac{br^2}{(b+r)^3}.$$

(3) 当 $c>0$, $d=0$ 时, 为传染病模型, 此时每次取出球后会增加下一次取到同色球的概率, 或者说, 每发现一个传染病患者, 以后都会增加再传染的概率. 同样的, 上述三种概率相等, 且都等于

$$P(B_1R_2R_3) = P(R_1B_2R_3) = P(R_1R_2B_3) = \frac{br(r+c)}{(b+r)(b+r+c)(b+r+2c)}.$$

可以看出, 当 $d=0$ 时, 只要取出的黑球和红球个数确定, 则概率不依赖于其抽出球的顺序.

(4) 当 $c=0$, $d>0$ 时, 为安全模型, 可以解释为, 每当事故发生, 会抓紧安全工作, 从而下一次发生事故的概率会减少, 而当事故未发生时, 安全工作会松懈, 下一次发生事故的概率会增大, 上述三种概率分别为:

$$P(B_1R_2R_3) = \frac{b}{(b+r)}\cdot\frac{r+d}{b+r+d}\cdot\frac{r+d}{b+r+2d},$$

$$P(R_1B_2R_3) = \frac{r}{(b+r)}\cdot\frac{b+d}{b+r+d}\cdot\frac{r+d}{b+r+2d},$$

$$P(R_1R_2B_3) = \frac{r}{(b+r)}\cdot\frac{r}{b+r+d}\cdot\frac{b+2d}{b+r+2d}.$$

例 1.4.3. 设 n 件产品中有 m 件不合格品, 从中任取两件, 已知两件中有一件是合格品, 求另一件也是合格品的概率.

解. 记事件 A 为"有一件是合格品", B"另一件也是合格品". 则

$$\begin{aligned}P(A) =&P(\text{取出一件合格品, 一件不合格品}) + P(\text{取出两件都是合格品})\\ =&\frac{C_m^1 C_{n-m}^1}{C_n^2} + \frac{C_{n-m}^2}{C_n^2} = \frac{2m(n-m)+(n-m)(n-m-1)}{n(n-1)}\\ =&\frac{(n-m)(n+m-1)}{n(n-1)}.\end{aligned}$$

因此 $P(AB) = P(\text{取出两件都是合格品}) = \frac{C_{n-m}^2}{C_n^2} = \frac{(n-m)(n-m-1)}{n(n-1)}$, 于是所求概率为

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{\frac{(n-m)(n-m-1)}{n(n-1)}}{\frac{(n-m)(n+m-1)}{n(n-1)}} = \frac{n-m-1}{n+m-1}.$$

□

1.4.2 事件的独立性

独立性是与条件概率密切相关的概率. 直觉上说, 事件 A 和 B 独立是说 A 的发生与否不影响 B 的发生与否, 反之亦然. 用概率的语言说, 就是 $P(A) = P(A|B)$ 且 $P(B) = P(B|A)$, 由条件概率的定义容易验证二者是等价的, 且都等价于下面的定义.

定义 1.4.2. 设 $(\Omega, \mathscr{F}, P)$ 为概率空间, 如果事件 $A, B \in \mathscr{F}$ 满足

$$P(AB) = P(A)P(B),$$

则称事件 A 和 B **相互独立** (简称**独立**).

命题 1.4.1. *独立事件有如下性质:*

(1) 若 A, B 独立, 且 $P(B) > 0$, 则

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

(2) 若 A, B 独立, 则 A 与 $\overline{B}$, $\overline{A}$ 与 B, $\overline{A}$ 与 $\overline{B}$ 亦独立.

(3) 零概率事件及其对立的事件与任意的事件都独立. 即条件概率等于无条件概率.

例 1.4.4. 袋中有 a 只黑球和 b 只白球, 令 A: "第一次摸到黑球", B: "第二次摸到黑球". 讨论 A 和 B 的独立性.

(1) 放回情形. 因为

$$P(A)=\frac{a}{a+b},\quad P(AB)=\frac{a^2}{(a+b)^2},\quad P(\overline{A}B)=\frac{ab}{(a+b)^2},$$

所以

$$P(B)=P(AB)+P(\overline{A}B)=\frac{a^2+ab}{(a+b)^2}=\frac{a}{a+b},$$

故 $P(A)P(B)=P(AB)$.

(2) 不放回情形. 易知

$$P(A)=P(B)=\frac{a}{a+b},\quad P(\overline{A}B)=\frac{ab}{(a+b)(a+b-1)},$$

故 $P(A)P(B)\neq P(AB)$.

事件的独立性可以进一步拓展到 n 个事件.

定义 1.4.3. 设 $\{A_k\}_{k\leqslant n}\subset\mathscr{F}$. 如果对任意的 $k\leqslant n$ 和 $1\leqslant i_1<i_2<\cdots<i_k\leqslant n$, 都有

$$P\left(\bigcap_{j=1}^{n}A_{i_j}\right)=\prod_{j=1}^{k}P(A_{i_j}),$$

则称 $A_1,A_2,\cdots,A_n$ **相互独立** (简称**独立**) .

注 1. 若 n 个事件 $A_1,A_2,\ldots,A_n$ 相互独立, 则它们两两独立 (即对于任意 $1\leqslant i,j\leqslant n$, 都有 A_i 和 A_j 独立) , 但是反之不对, 你能试着举出一个反例吗?

例 1.4.5 (伯恩斯坦反例)**.** 一个均匀的正四面体, 其第一、二、三面分别涂上红、黄、蓝三种颜色第四面同时涂上以上三种颜色. 以 A,B,C 分别表示投一次四面体出现红、黄、蓝颜色朝下的事件, 则

$$P(A)=P(B)=P(C)=\frac{1}{2},\quad P(AB)=P(BC)=P(AC)=\frac{1}{4}.$$

从而 A,B,C 两两独立, 但是,

$$P(ABC)=\frac{1}{4}\neq P(A)P(B)P(C).$$

独立性与概率计算: 设 $A_1,A_2,\cdots,A_n$ 相互独立, 则

$$P(A_1\cup A_2\cup\cdots\cup A_n)=1-\prod_{i=1}^{n}P(\overline{A_i}).$$

例 1.4.6. 设有某型号的高射炮, 每门炮 (发射一发) 击中敌机的概率为 0.6, 现在若干门炮同时发射 (每炮射一发) , 问: 若要以 99% 的把握击中来犯的一架敌机, 至少需要配置几门高射炮?

解. 设 n 是需要配置的高射炮的门数, 记 $A_i=$ "第 i 门炮击中敌机" $(i=1,\cdots,n)$, $A=$ "敌机被击中" . 由于 $A=\bigcup_{i=1}^n A_i$, 于是要找到 n, 使得

$$P(A)=P\left(\bigcup_{i=1}^n A_i\right)\geqslant 0.99.$$

由于 $P(A)=1-P(\overline{A})=1-P(\bigcup_{i=1}^n\overline{A_i})$, 且 $\overline{A_1},\cdots,\overline{A_n}$ 相互独立, 故

$$P(A)=1-P(\overline{A_1})\cdots P(\overline{A_n})=1-0.4^n.$$

为使不等式成立, 必须且只需 $1-0.4^n\geqslant 0.99$. 由此得

$$n\geqslant \lg 0.01/\lg 0.4=5.026.$$

故至少需配置 6 门高射炮方能以 99% 的把握击中敌机. □

例 1.4.7. 设 A, B, C 三事件相互独立, 证明 $A-B$ 与 C 独立.

证明. 因为

$$\begin{aligned}P((A-B)C)&=P(AC-BC)=P(AC)-P(ABC)\\&=P(A)P(C)-P(A)P(B)P(C)\\&=(P(A)-P(A)P(B))P(C)\\&=(P(A)-P(AB))P(C)=P(A-B)P(C).\end{aligned}$$

所以 $A-B$ 与 C 独立. □

1.5 全概率公式和贝叶斯公式

与条件概率有关的公式有三个, 除了上节的乘法公式外, 在就是将要介绍的全概率公式和贝叶斯公式. 这些公式的论证简单, 但应用非常丰富. 同学们需要通过具体的例子掌握并灵活运用这些公式.

1.5.1 全概率公式

分类讨论是数学中重要的思想, 这种思想也可以引入到概率问题的求解中. 分类讨论首先需要将一个复杂的问题分解为不重复不遗漏的若干种情况, 将类似的思想应用到概率求解种, 就是下面的定义.

定义 1.5.1 (完备事件组)**.** 若 $\{B_n\}_{n\geqslant 1} \subset \mathscr{F}$ 满足两两互斥且 $\sum\limits_{n=1}^{\infty} B_n = \Omega$, 则称 $\{B_n\}_{n\geqslant 1}$ 为**完备事件组**.

给出了分类讨论的若干种情况后, 需要对每一种情况逐一解答. 为了求出事件 A 的概率, 在给出了必然事件 Ω 的一个划分 $\{B_n\}_{n\geqslant 1}$ 后, 需要求解每种情况的概率 $P(AB_n) = P(B_n)P(A|B_n)$. 最终的结果可以用如下的公式求解.

定理 1.5.1 (全概率公式)**.** 设 $\{B_n\}_{n\geqslant 1}$ 为完备事件组, 则对任意事件 A 都有

$$P(A) = \sum_{n=1}^{\infty} P(B_n)P(A|B_n).$$

注 2. 在上式中, 若 $P(B_n) = 0$, 则规定 $P(B_n)P(A|B_n) = 0$.

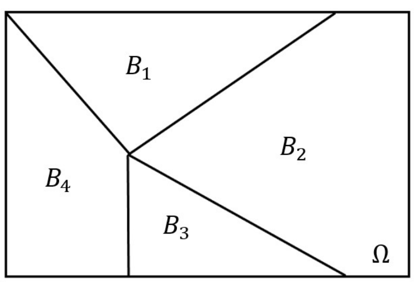

图 *1.2:* 完备事件组

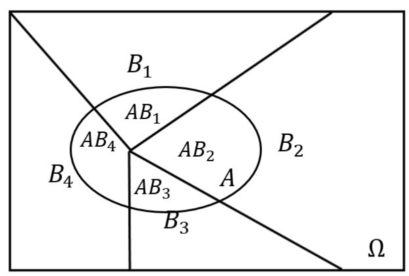

图 *1.3:* 全概率公式

例 1.5.1. 一保险公司相信人群可以分为 2 类: 一类是容易出事故的; 另一类是不容易出事故的. 已知前者在一年内出事故的概率为 0.4, 后者在一年内出事故的概率为 0.2. 前者约占人群的 30%. 今有一人前来投保, 他在一年内出事故的可能性有多大?

解. 设 $A =$ “他在一年内出事故”, $B =$ “他是容易出事故的”, 则 $B, \overline{B}$ 构成完备事件组, 有

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

由于 $P(B)=0.3, P(A|B)=0.4, P(\overline{B})=0.7, P(A|\overline{B})=0.2$, 于是

$$P(A)=0.3\times 0.4+0.7\times 0.2=0.26.$$

□

例 1.5.2. 甲口袋有 1 个黑球, 2 个白球, 乙口袋有 3 个白球, 每次从两口袋中任取一球, 交换后放入另一口袋中, 求交换 n 次之后, 黑球仍然在甲口袋的概率.

设事件 A_i 为“第 i 次交换后黑球仍然在甲口袋中”, 记 $p_i=P(A), i=0,1,2,\cdots$, 则有 $p_0=1$, 且

$$P(A_{i+1}\mid A_i)=\frac{2}{3},\quad P(A_{i+1}\mid A_i^c)=\frac{1}{3}.$$

由全概率公式得

$$p_n=\frac{2}{3}p_{n-1}+\frac{1}{3}(1-p_{n-1})=\frac{1}{3}p_{n-1}+\frac{1}{3},\quad n\geqslant 1,$$

得到递推公式

$$p_n-\frac{1}{2}=\left(\frac{1}{3}\right)\left(p_{n-1}-\frac{1}{2}\right),\quad n\geqslant 1.$$

将 $p_0=1$ 代入上式可得

$$p_n-\frac{1}{2}=\left(\frac{1}{3}\right)^n\left(\frac{1}{2}\right),$$

因此

$$p_n=\frac{1}{2}\left[1+\left(\frac{1}{3}\right)^n\right].$$

1.5.2 贝叶斯公式

贝叶斯公式的主要目的是通过 $P(A|B)$ 反过来求解 $P(B|A)$.

定理 1.5.2 (贝叶斯公式)**.** 设 $\{B_n\}_{n\geqslant 1}$ 为完备事件组, $A\in\mathscr{F}$ 满足 $P(A)>0$, 则

$$P(B_n|A)=\frac{P(B_n)P(A|B_n)}{\sum\limits_{n=1}^{\infty}P(B_n)P(A|B_n)}.\tag{1.5.1}$$

注 3. 在使用贝叶斯公式 (1.5.1) 时, 常将 $P(B_i)$ 叫做**先验概率**, 而将在 A 发生的条件下 B_i 发生的概率 $P(B_i|A)$ 称为**后验概率**. 因此贝叶斯公式 (1.5.1) 可以看作从先验概率到后验概率的转换公式.

注 4. 贝叶斯公式描述了这样的一个过程. 我们关心的是 $\{B_n\}_{n\geqslant 1}$ 发生的概率. 在最初没有任何“信息”, 因此是先验概率 $P(\cdot)$. 之后我们知道了 A 发生的“信息”, 则概率“修正”为后验概率 $P(\cdot|A)$ (条件概率也是概率) . 贝叶斯公式提供了具体的计算方法.

例 1.5.3. 一项血液化验有 95% 的把握诊断某种疾病, 但这项化验用于健康人也会有 1% 的"假阳性"结果 (即如果一个健康人接受这项化验, 化验结果误诊此病人患该疾病的概率为 1%) . 假定该疾病的患者事实上只占总人口的 0.5%. 若某人化验结果为阳性, 则此人确实患有该疾病的概率是多少?

解. 令 A 表示"此人确实患该疾病", B 表示"其化验结果为阳性", 则所求概率为

$$\begin{aligned} P(A|B) =& \frac{P(A)P(B|A)}{P(A)P(B|A)+P(\overline{A})P(B|\overline{A})} \\ =& \frac{0.95\times 0.005}{0.95\times 0.005+0.01\times 0.995} = \frac{95}{294}. \end{aligned}$$

□

例 1.5.4. 某厂有甲、乙、丙三个车间生产同一种产品. 已知这三个车间的产量分别占总产量的 25%, 35%, 40%. 不合格品率分别是 0.05, 0.04, 0.02. 有一用户买了该厂一件产品, 经检验是不合格品. 但该产品是哪个车间生产的标志已脱落. 问: 该产品分别是车间甲、乙、丙生产的概率是多少?

解. 设 A = "从该厂产品中任取一件恰好取到不合格品", B_1 = "该产品是车间甲生产的", B_2 = "该产品是车间乙生产的", B_3 = "该产品是车间丙生产的". 要求的是条件概率 $P(B_1|A)$, $P(B_2|A)$, $P(B_3|A)$. 已知 $P(B_1) = 25/100$, $P(B_2) = 35/100$, $P(B_3) = 40/100$, $P(A|B_1) = 0.05$, $P(A|B_2) = 0.04$, $P(A|B_3) = 0.02$. 利用贝叶斯公式 (1.5.1), 得

$$P(B_1|A) = \frac{P(B_1)P(A|B_1)}{\sum_{i=1}^{3} P(B_i)P(A|B_i)} = \frac{25}{69}.$$

用同样的方法可计算得 $P(B_2|A) = \frac{28}{69}$, $P(B_3|A) = \frac{16}{69}$, 可见, 该产品来自乙车间的可能性最大. □

注 5. 上面的计算过程可以用图 1.4 表示, 箭头上标出来转移概率, 条件概率可以看成对 AB_i 发生的概率重新归一化.

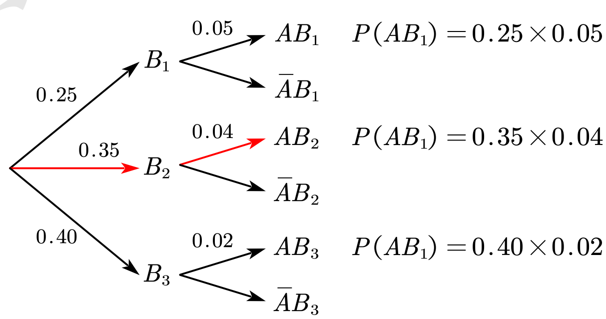

图 1.4: 图解贝叶斯公式.

例 1.5.5. 一架飞机失踪了, 推测它等可能的坠落在 3 个区域. 令 $\alpha_i(i=1,2,3)$ 表示飞机在第 i 个区域坠落但没有被发现的概率. 已知对区域 1 的搜索没有发现飞机, 求在此条件下, 飞机坠落在第 $i(i=1,2,3)$ 个区域的条件概率.

解. 令 B_i 表示"飞机坠毁在第 i 个区域", $i=1,2,3$, A 表示"在第 1 个区域没有搜索到飞机", 则

$$P(B_1|A)=\frac{P(B_1)P(A|B_1)}{\sum\limits_{i=1}^{3}P(B_i)P(A|B_i)}=\frac{\frac{\alpha_1}{3}}{\frac{\alpha_1}{3}+1\times\frac{1}{3}+1\times\frac{1}{3}}=\frac{\alpha_1}{\alpha_1+2}.$$

对 $j=2,3$,

$$P(B_j|A)=\frac{P(B_j)P(A|B_j)}{\sum\limits_{i=1}^{3}P(B_i)P(A|B_i)}=\frac{\frac{1}{3}}{\frac{\alpha_1}{3}+1\times\frac{1}{3}+1\times\frac{1}{3}}=\frac{1}{\alpha_1+2}.$$

□

随机游走: 考虑数轴上一质点, 假定它只在整数点上运动. 当前时刻它处于位置 a (整数), 下一时刻 (单位间隔时间) 以概率 p 向正向, 概率 $1-p$ 向负向运动一个单位, 称这样的质点运动为随机游动, 当 $p=q=\frac{1}{2}$ 时, 称为对称随机游走.

(1) 无限制随机游走: 对随机游走, 以 S_n 表示 n 时刻质点的位置, 假定 $S_0=0$. 我们计算经过 n 次运动后到达位置 k 的概率.

由于质点在 n 时刻位于 k, 在 n 次游动中, 质点向右移动次数 x 比向左运动 y 多 k 次:

$$x-y=k,\quad x+y=n,$$

$$x=\frac{n+k}{2},\quad y=\frac{n-k}{2}.$$

为使 x 为整数, k 和 n 的奇偶性需要相同, 即

$$P(S_n=k)=\begin{cases}C_n^{\frac{n+k}{2}}p^{\frac{n+k}{2}}(1-p)^{\frac{n-k}{2}}, & n,k\text{ 奇偶性相同;}\\ 0, & n,k\text{ 奇偶性不同.}\end{cases}$$

(2) 两端带有吸收壁的随机游走: 设 a, b 为正整数. 假定质点初始位置为 a, 在位置 0 和 $a+b$ 均有一个吸收壁, 求质点被吸收的概率.

记 q_n 为质点初始位置是 n 而最终在 $a+b$ 被吸收的概率, 显然,

$$q_0=0,\quad q_{a+b}=1.$$

若质点某时刻位于 $n, n=1,\cdots,a+b-1$. 则其在位置 $a+b$ 被吸收有两种可能: (1) 运动到 $n-1$ 位置被 $a+b$ 吸收, (2) 运动到 $n+1$ 位置被 $a+b$ 吸收, 由全概率公式得

$$q_n = q_{n-1}q + q_{n+1}p, \quad n=1,\cdots,a+b-1.$$

由于 $p+q=1$, 上式可以写为

$$p(q_{n+1}-q_n) = q(q_n - q_{n-1}), \quad n=1,\cdots,a+b-1.$$

记 $r=\frac{q}{p}$, 则

$$q_{n+1}-q_n = r(q_n - q_{n-1}), \quad n=1,\cdots,a+b-1.$$

可以分两种情况讨论: (i) 若 $r=1$, 即 $p=q=\frac{1}{2}$. 则

$$q_{n+1}-q_n = q_n - q_{n-1} = \cdots = q_1 - q_0,$$

$$q_{n+1} = q_0 + (n+1)(q_1-q_0), \quad n=1,\cdots a+b-1.$$

结合边值条件, 有

$$q_n = \frac{n}{a+b}, n=1,\cdots,a+b-1.$$

(ii) 若 $r\neq 1$, 即 $p\neq q$:

$$q_{n+1}-q_n = r(q_n-q_{n-1}) = \cdots = r^n(q_1-q_0),$$

即

$$q_n - q_0 = \sum_{i=0}^{n-1}(q_{i+1}-q_i) = \sum_{i=0}^{n-1} r^i(q_1-q_0) = \frac{1-r^n}{1-r}(q_1-q_0), \quad n=1,\cdots,a+b-1.$$

结合边值条件, 得

$$q_1 = \frac{1-r}{1-r^{a+b}},$$

则

$$q_n = \frac{1-r^n}{1-r^{a+b}}.$$

1.6 本章总结

本节主要的知识点可分为 3 个模块：概率空间的定义、古典概型、条件概率与贝叶斯公式.

1. **概率空间的定义.** 概率空间的数学定义以集合论作为基础. 首先要注意概率空间的定义源于一个“上帝”视角, 实验的结果被称为样本点, 所有样本点的集合被称为样本空间. 事件被定义为样本空间的子集. 这样事件的交与并由集合论的交与并决定 (回忆画图法) .

   概率的数学定义是一个函数, 它将事件映射成发生的可能性 (0 到 1 之间的数) . 概率的几何直觉是集合的面积, 其中重点掌握全空间的概率为 1 与事件交并后的概率求法 (推论 1.1.1) .

2. **古典概型.** 古典概型是一种基础的概率模型, 在这类问题中存在基础事件, 基础事件概率相同, 待求的事件由若干基础事件构成. 所以待求事件的概率等于构成它的基础事件数除以总事件数 (见式 (1.3.1)) . 古典概型的问题被转化成高中排列组合的问题. 所以古典概型问题求解主要有两个步骤：1、明确基础事件 2、计算总基础事件数与待求事件的基础事件数。考试难度参考例 1.3.1 和例 1.3.2.

3. **条件概率与贝叶斯公式.** 条件概率研究在某个条件下, 事件发生的概率. 掌握定义 1.4.1. 数学直觉上, 条件概率 $P(A|B)$ 是将 B 看成全集, 求 A 和 B 的交空间占用 B 的面积. 掌握例 1.4.1 和例 1.4.2. 两个事件独立是说两个事件共同发生的概率等于这两个事件概率的乘积 (掌握定义 1.4.2 与定义 1.4.3) . 直觉上, 如果两个事件没有什么关系, 其中一个事件发生对另一个事件发生不会有影响, 即 $P(A) = P(A|B)$. 掌握例 1.4.4 与 1.4.6.

   贝叶斯公式重点理解先验概率与后验概率含义. 掌握画图法求解后验概率问题 (见例 1.5.4 和图 1.4) , 掌握例 1.5.3.

1.7 有趣的故事

1.7.1 概率论的发展简史

概率论的故事始于人类对不确定性的古老兴趣, 其最初的动力并非来自高深的科学, 而是源于一种古老的概率游戏——赌博. 尽管历史悠久的活动表明人们早已有量化“可能性”的动机, 但将其转化为精确的数学语言, 却是一条漫长的道路.

16 世纪, 意大利博学家吉罗拉莫 · 卡尔达诺 (Gerolamo Cardano) 首次尝试用数学方式思考胜率, 他将“几率”(odds) 定义为有利与不利结果之比, 这已隐含了现代概率的雏形. 然而, 真正的奠基时刻发生在 17 世纪. 1654 年, 布莱兹 · 帕斯卡 (Blaise Pascal) 和皮埃尔 · 德 · 费马 (Pierre de Fermat) 这两位数学巨匠通过一系列著名的通信, 系统地解决了赌博中遇到的分配问题, 这被后世公认为数学概率论的开端. 随后, 克里斯蒂安 · 惠更斯 (Christiaan Huygens) 于 1657 年发表了第一篇关于概率的正式论文, 开始了这门学科的科学化进程.

19 世纪的概率论的研究不再局限于赌桌. 高斯和勒让德等人将其广泛应用于天文学和测量误差分析中, 发展出了诸如最小二乘法和正态分布等关键工具, 使概率论成为自然科学的重要基石. 进入 20 世纪, 概率论迎来了其成熟的关键标志——公理化. 1933 年, 俄罗斯数学家安德雷 · 柯尔莫哥洛夫 (Andrey Kolmogorov) 在其著作中运用了测度论, 为概率论建立了坚实的公理体系. 在这一框架下, 概率被定义为一个满足三条基本公理的“测度”, 所有复杂的概率问题都可以从这个严格的数学基础中推导出来. 几乎在同一时代, Cox 则从逻辑命题的角度出发, 给出了另一种等价的表述.

至此, 概率论彻底摆脱了早期的直观与经验色彩, 蜕变为一门结构严谨、逻辑严密的现代数学分支. 它的诞生虽与赌博游戏息息相关, 但其发展却远远超越了游戏, 成为了我们理解和描述不确定世界的核心语言.

1.7.2 人工智能的发展

1956 年的达特茅斯会议首次明确提出了“人工智能”这一研究方向, 成为后来一切探索的起点. 在会后的几年里, 符号主义方法迎来了“大发现的时代”. 研究者们开发出了一系列令人惊叹的程序：它们能够解决代数应用题、证明几何定理, 甚至学习和使用自然语言. 在当时, 人们几乎难以相信计算机竟能展现出如此“智能”的行为. 学界在 1970 年代后转向专家系统. 专家系统依靠知识库和推理规则, 能够在特定领域中表现出接近专家的水平. XCON 等系统曾大获成功, 证明了人工智能的应用潜力. 然而, 它们也存在维护成本高、难以升级、对异常情况极其脆弱等问题, 最终难以扩展到通用智能. 尽管符号主义的方法清晰, 在初期取得了成功, 但在辉煌之后逐渐显露出根本性瓶颈. 在当时, 计算机有限的内存和处理速度不足以解决任何实际的 AI 问题. 然而这一问题并不能随着硬件的发展从而得到解决, 这是因为很多问题的计算复杂度数量是指数爆炸的. 例如, 判断一个布尔方程式是否存在解 (SAT 问题) 是一个 NPC 问题, 基本排除了在多项式时间内求解的可能.

符号主义方法的另一个困难对“不确定性”建模. 回归模型是一个典型的进行不确定性的预测的统计方法. 在 1980 年代, 统计方法逐渐进入人工智能的主流研究, 贝叶斯方法和隐马尔可夫模型等统计机器学习模型, 被广泛应用于语音识别和自然语言处理. 1990 年代, Vapnik 的统计学习理论为机器学习提供了严格的数学框架. 与此同时, 支持向量机和核方法作为机器学习的代表性成果, 极大地推动了 AI 的发展. 2000 年前后的概率图模型 (PGM) (如贝叶斯网络、条件随机场) 是对复杂数据建模的概率统计工具. 这一阶段 AI 从确定性逻辑走向概率建模, AI 的核心越来越多地依赖于统计建模方法.

随着计算硬件的发展, 算力的提高让连接主义再次得到应用. 2012 年, AlexNet 在 ImageNet 竞赛中大幅领先, 标志着深度学习的兴起. 深度学习采用的神经网络是一种数据驱动的方法, 不再依赖精确的推理. 连接主义的兴起并非偶然：一方面, GPU 等硬件算力的飞跃让训练庞大神经网络成为可能; 另一方面, 随机梯度下降等高效的随机化方法有效降低了计算成本. 硬件与算法的进步, 让人工智能从象牙塔走向现实应用, 开启了新的快速发展阶段.

1.7.3 贝叶斯公式和贝叶斯学派

贝叶斯公式是概率论中最重要的定理之一, 用来描述在已有证据下如何更新对某一事件发生概率的认识. 它最早由英国牧师、统计学家托马斯 · 贝叶斯 (Thomas Bayes, 1702–1761) 提出. 贝叶斯在其遗稿“An Essay Towards Solving a Problem in the Doctrine of Chances”中, 研究了如何根

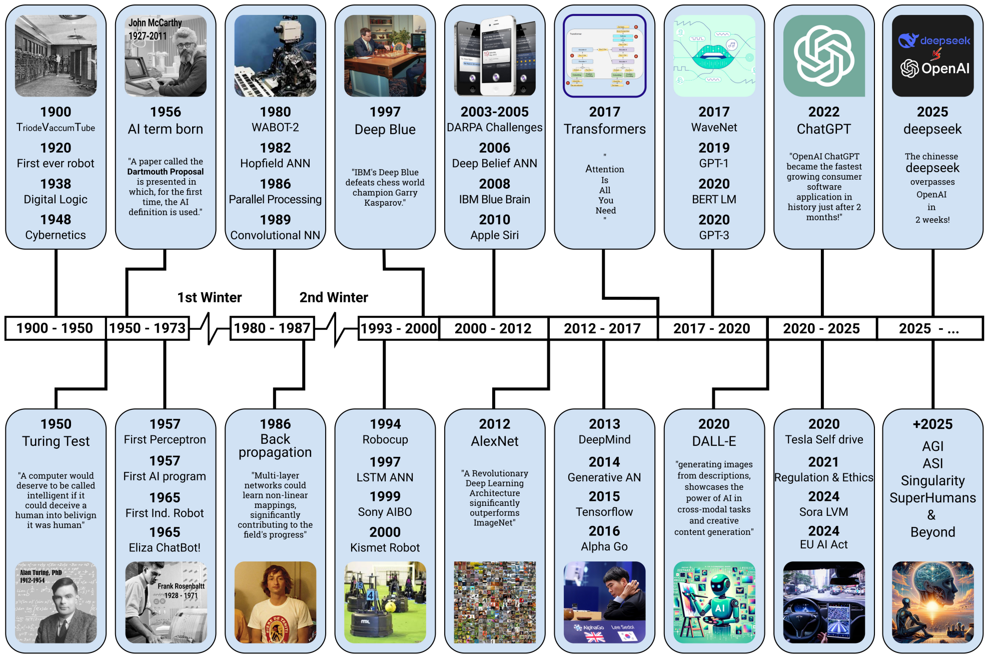

图 1.5: 人工智能发展的时间线.

据实验结果来推断二项分布中未知参数的分布, 这实际上就是现代所谓的“贝叶斯推断”的雏形.

贝叶斯去世后, 他的朋友理查德 · 普赖斯 (Richard Price) 发现并整理了这份手稿, 并在 1763 年提交给英国皇家学会公开发表. 普赖斯不仅修改并完善了文章, 还在引言中阐述了其哲学意义, 并将方法应用于人口与寿命年金的计算. 到了 20 世纪, 哈罗德 · 杰弗里斯 (Harold Jeffreys) 为贝叶斯方法奠定了公理化的基础, 他曾形容“贝叶斯定理之于概率论, 就如同毕达哥拉斯定理之于几何学”.

贝叶斯公式不仅是一个计算条件概率的定理, 而且衍生出了一整套方法论, 在统计学中开创了与频率学派平行的贝叶斯学派. 其核心思想是: 把需要估计未知参数本身也看作随机变量, 通过先验分布 (prior) 来表达我们在观测数据之前对未知参数的信念, 再利用贝叶斯公式结合数据更新得到后验分布 (posterior). 这种方法强调“随着信息增加, 认知不断修正”. 在进入 21 世纪后, 随着计算能力的提升, 贝叶斯方法在机器学习、人工智能中有了广泛的应用.