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、张海涵对讲义整理的帮助。

1.1 随机事件及其运算

1.1.1 随机事件

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 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

- 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。

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

随机现象的某些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, E 为抛掷一枚骰子, 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

事件的频率：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 $\frac{\mu}{n}$，随着实验次数 n 增大，频率会在某一数值 p 附近摆动，称为该事件的**概率**，记为 $P(A) = p$

由于频率 $\frac{\mu}{n}$ 总在 $0, 1$ 之间，我们有：

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

例如投一枚硬币 n 次，出现 μ 次正面，则 $\frac{\mu}{n} \overset{n\to\infty}{\to} p$。其中，主观概率 p 为事件的置信度，概率是可能性大小的度量。大概率事情易发生，小概率事情不易发生。

1.1.1.1 事件的交和并

定义 2.1 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

$$A \subset B\ (\text{或 } B \supset A)$$

定义 2.2 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

$$A = B$$

定义 2.3 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 2.1 （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 $C = A \cup B$.

对于并运算，有以下性质，我们恒记必然事件为 U，不可能事件为 V：

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

定义 2.4 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 2.1 中，$A \cap C = A$，$B \cap C = C$，$A \cap B = A$

对于交运算，有以下性质：

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

1.1.1.2 事件的余和差

定义 2.5 设 A 是事件，称 "非 A" 是 A 的对立事件（或称余是事件），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 2.6 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 $A \backslash B$）

由定义知，$A - B = A \cap \bar{B}$
画图法确定关系。

1.1.1.3 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

多个事件的交和并：
设 $A_1, A_2, \cdots, A_n$ 是 n 个事件，则“$A_1, A_2, \cdots, A_n$”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，常常用 $\bigcup\limits_{i=1}^{n} A_i$ 表示 $A_1, A_2, \cdots, A_n$ 的并
设 $A_1, A_2, \cdots, A_n$ 是 n 个事件，则“$A_1, A_2, \cdots, A_n$”的交是指这样的事件：它发生当且仅当 $A_1, A_2, \cdots, A_n$ 这 n 个事件都发生，常常用 $\bigcap\limits_{i=1}^{n} A_i$ 表示 $A_1, A_2, \cdots, A_n$ 的交，也用 $A_1 A_2 \cdots A_n$ 表示这个“交”

实际应用中，还需定义无穷多事件的并与交
设 $A_1, A_2, \cdots, A_i, \cdots$ 是一列事件，则 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，这个 B 叫做诸 A_i 的并，记为 $\bigcup\limits_{i=1}^{\infty} A_i$，有时也写为 $A_1 \cup A_2 \cup \cdots$
设 $A_1, A_2, \cdots, A_i, \cdots$ 是一列事件，则 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，这个 C 叫做诸 A_i 的交，记为 $\bigcap\limits_{i=1}^{\infty} A_i$，有时也写为 $A_1 A_2 \cdots$

例: 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 $\bigcup\limits_{i=1}^{n} A_i$ 发生等价于 $X \in [\frac{1}{n+1}, 1]$，事件 $\bigcap\limits_{i=1}^{n} B_i$ 发生等价于 $X \in [0, \frac{1}{n}]$。进而当 $n \to \infty$ 时事件 $\bigcup\limits_{i=1}^{\infty} A_i$ 发生等价于 $X \in (0, 1]$，事件 $\bigcap\limits_{i=1}^{\infty} B_i$ 发生等价于 $X = 0$。

并的更一般定义是，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），这些事件 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，这个 B 常常记为 $\bigcup\limits_{a \in \Gamma} A_a$，类似可以定义一族事件的交 $\bigcap\limits_{a \in \Gamma} A_a$

例 2.3:（对应郑书例 2.3）一射手向一个目标连续射击，设 $A_1 =$ "第一次射击，命中"，$A_i =$ "前 $i-1$ 次射击都未命中，第 i 次射击命中"（$i = 2, 3, \cdots$），$B =$ "终于命中"，则 $B = \bigcup\limits_{i=1}^{\infty} A_i$

例 2.4:（对应郑书例 2.4）一射手向一个目标连续射击，设 $A_i =$ "第 i 次射击，未命中目标"（$i = 2, 3, \cdots$）则 $\bigcap\limits_{i=1}^{\infty} A_i =$ "每次均未命中目标"

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

1.1.1.4 互斥事件

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）

称事件 $A_1, \cdots A_n$ 互不相容，若对任何 $i \neq j (i, j = 1, \cdots n)$，$A_i$ 与 A_j 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

- 加法公式：$A_1, A_2, \ldots$ 互不相容，则：

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

- $P(A \bigcup B) = P(A) + P(B) - P(AB)$

1.2 概率的公理化定义

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

例:

- $\mathscr{F} = \{\emptyset, \Omega\}$ 平凡概率空间子类
- $\mathscr{F} = \{\emptyset, \Omega, A, \overline{A}\}$ 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: 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:

- 非负性：$P(A) \geqslant 0$ 对于一切 $A \in \mathscr{F}$
- 规范性：$P(\Omega) = 1$
- 可列可加性：若 $A_n \in \mathscr{F}(n = 1, 2, \cdots)$ 两两不相交, 则

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

$(\Omega, \mathscr{F}, P)$ 为概率空间

例 1: 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

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

则 $\mathbb{P}$ 为 $(\Omega, \mathscr{F})$ 上概率。
概率 P 有以下性质:
(1) $P(\emptyset) = 0$;
(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

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

2.1 古典概型

模型定义：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

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

$$P(\{w_1\}) = \cdots = P(\{w_n\})$$

则 $(\Omega, \mathscr{F}\ P)$ 为古典概型对应的概率空间。
计算公式：对 $A = \{\omega_{i_1}, \cdots, \omega_{i_k}\} \in \mathscr{F}$，利用概率的有限可加性可知：

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

排列：从含有 n 个不同元素的总体中抽取 r 个进行排列
(1) 放回情形：共有 n^r 种排列方式
(2) 不放回情形：共有 $A_n^r := n(n-1)\cdots(n-r+1)$ 种排列方式
当 $r = n$ 时，为全排列，此时 $A_n^n = n!$。
组合：(1) 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 $C_n^r = \frac{n!}{r!(n-r)!} = \frac{A_n^r}{r!}$
(2) 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
(3) 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 $C_{n_1}^{r_1} C_{n_2}^{r_2} \cdots C_{n_k}^{r_k}$。
(4) 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：（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 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 $C_8^1C_1^10$ 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 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 $\frac{C_{2n-r}^n}{2^{2n-r}}$.

2.2 条件概率与独立性

2.2.1 条件概率

条件概率： 设 $(\Omega, \mathscr{F}\ P)$ 为概率空间，$B \in \mathscr{F}$ 满足 $P(B) > 0$。称

$$P(A|B) = \frac{P(AB)}{P(B)}, A \in \mathscr{F}$$

为 B 发生条件下 A 发生的条件概率。

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

- $P(A|B) \geqslant 0, \quad \forall A \in \mathscr{F}$
- $P(\Omega|B) = 1$
- $\forall \{A_n\} \subset \mathscr{F}, A_n \cap A_m = \emptyset, \forall n \neq m,$

$$P(\sum_{n=1}^{\infty} A_n|B) = \sum_{n=1}^{\infty} P(A_n|B)$$

容易得到，$P(B|\Omega) = P(B)$。

乘法公式： $P(AB) = P(B \mid A)P(A)$

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 $P(A_1A_2\cdots A_{n-1} > 0$。

例 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 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： 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 "有一件是合格品"，B "另一件也是合格品"。则

$P(A)=P$（取出一件合格品，一件不合格品）$+P$（取出两件都是合格品）

$=\frac{C_m^1C_{n-m}^1}{C_n^2}+\frac{C_{n-m}^2}{C_n^2}=\frac{2m(n-m)+(n-m)(n-m-1)}{n(n-1)}$

$=\frac{(n-m)(n+m-1)}{n(n-1)}$

$P(AB)=P$（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

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

2.2.2 事件的独立性

事件的独立性： 设 $(\Omega,\mathscr{F}\ P)$ 为概率空间，称 $A,B\in\mathscr{F}$ 相互独立（独立），若

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

性质：（1）若 A,B 独立，且 $P(B)>0$，则

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

即条件概率等于无条件概率。

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

$$P(A\overline{B})=P(A)-P(AB)=P(A)-P(A)P(B)=P(A)P(\overline{B})$$

（3）零概率事件及其对立的事件与任意的事件都独立。

例： 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

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

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

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

故

$$P(A)P(B)\neq P(AB)$$

定义： 设 $\{A_k\}_{k\leqslant n}\subset\mathcal{F}$。称 $A_1,A_2,\cdots,A_n$ 相互独立，若

$$P(\bigcap_{j=1}^{k}A_{i_j})=\prod_{j=1}^{k}P(A_{i_j}),\quad 1\leqslant i_1<i_2<\cdots<i_k\leqslant n,k\leqslant n$$

注意：独立 $\Rightarrow$ 两两独立，但是反之不对：
伯恩斯坦反列：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： 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

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

由于 $P(A)=1-P(\overline{A})=1-P(\overset{n}{\underset{i=1}{\cup}}\overline{A_i})$，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： 设 A，B，C 三事件相互独立，证明 $A-B$ 与 C 独立。

解： 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。 □

2.3 全概率公式和贝叶斯公式

2.3.1 全概率公式

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。

全概率公式： 假定 $\{B_n\}_{n\geqslant 1}$ 为完备事件组，则

$$P(A)=\sum_{n=1}^{\infty}P(B_n)P(A|B_n),\forall A\in\mathscr{F}$$

注意：在上式中，若 $P(B_n)=0$，则规定 $P(B_n)P(A|B_n)=0$。

例： 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

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

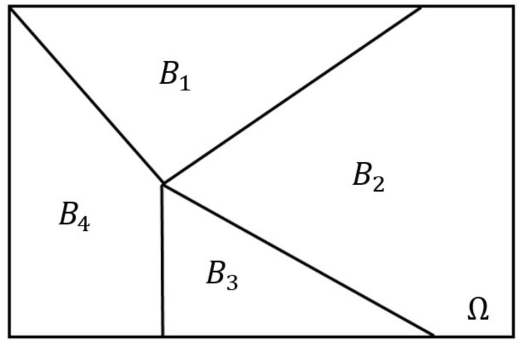

图 *2.1:* 完备事件组

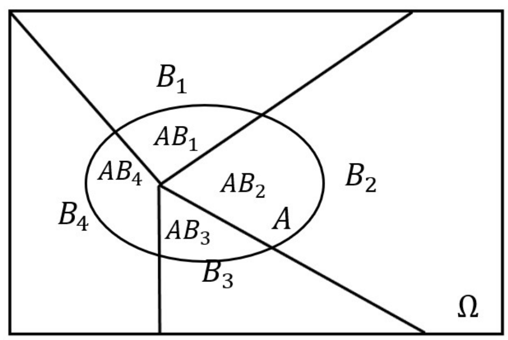

图 *2.2:* 全概率公式

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

$$P(A)=0.3\times0.4+0.7\times0.2=0.26$$

$\square$

例： 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

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

□

2.3.2 贝叶斯公式

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

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

例：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

$$\begin{aligned}P(A|B)&=\frac{P(A)P(B|A)}{P(A)P(B|A)+P(\overline{A})P(B|\overline{A})}\\&=\frac{0.95\times 0.005}{0.95\times 0.005+0.01\times 0.995}\\&=\frac{95}{294}\approx 0.323\end{aligned}$$

□ **例：**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

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

3.1 随机变量

为了进一步研究随机现象，我们需要引入随机变量的概念。

定义：（随机变量的直观描述）如果条件 S 下的结果可以用某个变量 X 来描述，X 的值不能预先确定，而随着条件 S 的不同可能变化，但是对任何实数 c，事件“X 取值不超过 c”是有概率的，将这样一种变量 X 称为随机变量。

定义：（随机变量的数学描述）如果条件 S 下的所有可能结果组成了集合 $\Omega = \{\omega\}$，$X = X(\omega)$ 是在 Ω 上有定义的实值函数，而且对任何实数 c，事件“$\{\omega : X(\omega) \leqslant c\}$”是有概率的，将 X 称为随机变量。

例:（对应郑书例 1.2）盒中有 5 个球, 其中有 2 个白球, 3 个黑球. 从中任取 3 个球, 将其中所含的白球的数目记为 X.

建模: 将球编号, $1 \sim 3$ 表示黑球, 4,5 表示白球.

记摸到球的编号为 $\omega = (i, j, k)$, 其中 $1 \leqslant i < j < k \leqslant 5. |\Omega| = C_5^3 = 10$.

其中满足 $X = 0$ 的 ω 有 $C_2^0 C_3^3 = 1$ 个;　满足 $X = 1$ 的 ω 有 $C_2^1 C_3^2 = 6$ 个;　满足 $X = 2$ 的 ω 有 $C_2^2 C_3^1 = 3$ 个.

设事件: $\{X = 1\} = \{\omega : X(\omega) = 1\}, \quad \{X \leqslant 1\} = \{\omega : X(\omega) \leqslant 1\}$.

将 $P(\{X = 1\})$ 简记为 $P(X = 1)$.

$$P(X = 1) = \frac{6}{10},\ P(X \leqslant 1) = \frac{7}{10}.$$

例:（对应郑书例 1.6）某公共汽车站每隔 10 min 会有一两某路公交车到达. 某乘客随机在任意时刻到达车站.

显然，他的候车时间 X (单位: min) 为随机变量. X 的取值范围 $0 \leqslant X \leqslant 10$。事件 $\{X \leqslant c\}$ 是有概率的，这是一种几何概型，我们会在后面给出计算过程，例如:

$$P(X \leqslant 3) = \frac{3}{10}, \quad P(2 \leqslant X \leqslant 6) = \frac{4}{10}.$$

3.2 离散型随机变量

定义： X 是离散型随机变量指: X 取有限个值 $x_1, \cdots, x_n$, 或可列无穷个值 $x_1, x_2, \cdots$.X 的概率分布 (列) 指:

$$p_k = P(X = x_k), \quad k = 1, \cdots, n \text{ 或} k = 1, 2, \cdots.$$

将 X 的可能值以及相应的概率列为表3.1。表3.1称为 X 的**概率分布表**，它能够清楚完整的表示 X

X	x_1	x_2	$\cdots$	x_k	$\cdots$
p	p_1	p_2	$\cdots$	p_k	$\cdots$

表 3.1: 概率分布表

的取值以及概率的分布情况。

定义： 设 X 的可能取值是 $x_1, x_2, \cdots$（有限个或者可列无穷个），则称

$$p_k = P(X = x_k) \quad (k = 1, 2, \cdots)$$

为 X 的**概率分布**，这时也称为 X 的**概率函数**或者**概率分布律**

关于 $\{p_k\}$，有以下性质：

（1）$p_k \geqslant 0 \ (k = 1, 2, \cdots)$ （2）$\sum\limits_k p_k = 1$

回忆本讲例 1 的 X（抽到的白球数）它的概率分布表如表3.2所示：

X	0	1	2
p	0.1	0.6	0.3

表 3.2: X 的概率分布表

对离散型随机变量，有以下几种常见的概率分布：

3.2.1 两点分布（伯努利分布）

定义随机变量 X 的可能值是 0 和 1 且概率分布为:

$$P(X=1)=p,\quad P(X=0)=1-p.$$

称 X 服从**两点分布**（也称伯努利分布），记为 $X\sim B(1,p)$ (参数 $0\leqslant p\leqslant 1$)

我们定义示性函数 1_A：事件 A 发生则取 1; A 不发生则取 0 .

例:（对应郑书例 2.1）100 件产品中有 3 件次品. 从中任取一件.

设事件 $A=$ “取到合格品”，, 随机变量 $X=1_A$，X 的可能取值为 0 和 1。取到每件产品的概率均等，概率分布为

$$P(X=1)=\frac{97}{100},P(X=0)=\frac{3}{100}$$

X 服从参数 $p=0.97$ 的两点分布。

3.2.2 二项分布

设随机变量所有可能值为 $0,1,\cdots,n$，且

$$P(X=k)=C_n^kp^k(1-p)^{n-k},k=0,1,\cdots,n$$

称 X 服从参数为 n,p 的二项分布，记作 $X\sim B(n,p)$(参数 $n\geqslant 1,0\leqslant p\leqslant 1$)

二项分布有明显的实际背景，例如在单次实验中事件 A 发生的概率是 p，进行独立重复实验 n 次，记事件 A 发生的次数为 X，则 $X\sim B(n,p)$。

定理 2.1: 对于二项分布，分布列 $P(X=k)$ 的最大值点 k_0 如下:

若 $(n+1)p\notin\mathbb{Z}$, 则 $k_0=[(n+1)p]$;

若 $(n+1)p\in\mathbb{Z}$, 则 $k_0=(n+1)p$ 或 $(n+1)p-1$.

证明：显然

$$\frac{p_n(k+1)}{p_n(k)}=\frac{n-k}{k+1}\cdot\frac{p}{1-p}$$

由于 $\frac{n-k}{k+1}\cdot\frac{p}{1-p}>1$ 等价于 $k<(n+1)p-1$，于是有:

(a) 当 $k<(n+1)p-1$ 时，$p_n(k+1)>p_n(k)$

(b) 当 $k>(n+1)p-1$ 时，$p_n(k+1)<p_n(k)$

(c) 当 $k=(n+1)p-1$ 时，$p_n(k+1)=p_n(k)$

(i) 若 $(n+1)p \notin \mathbb{Z}$，设 $k_0 = [(n+1)p] < (n+1)p < k_0+1$，当 $k < m$ 时，$k \leqslant k_0 - 1 < (n+1)p - 1$，因此 $p_n(k) < p_n(k+1)$；当 $k \geqslant k_0$ 时，$k > (n+1)p - 1$，因此 $p_n(k) > p_n(k+1)$，所以 k_0 为最大值。

(ii) 若 $(n+1)p \in \mathbb{Z}$，设 $k_0 = (n+1)p$，有 $p_n(k_0) = p_n(k_0+1)$，进而利用性质（a）和性质（b）知 k_0 为最大值。 □

3.2.3 泊松分布

定义： 设随机变量 X 的所有可能取值是全体非负整数，且

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \cdots.$$

则称 X 服从参数为 λ 的泊松分布，记为 $X \sim \mathscr{P}(\lambda)$(参数: $\lambda > 0$) 。

泊松分布常见于生物学，物理学，工业的应用中，例如电话交换台收到的电话呼唤次数，放射性物质在一定时间内放出的粒子数。

定理： 泊松分布的分布列最大值点 $k_0 = [\lambda]$。

证明：注意到 $p_{k+1} = \frac{\lambda}{k+1} p_k$，故由分布函数知

若 $k+1 \leqslant \lambda$，则 $p_{k+1} \geqslant p_k$

若 $k+1 \geqslant \lambda$，则 $p_{k+1} \leqslant p_k$

因此当 $k_0 = [\lambda]$ 时，分布列取最大值。 □

例： 已知某商场一天来的顾客服从参数为 λ 的泊松分布，而每个来商场的顾客购物概率为 p，证明此商场一天内购物的顾客数服从参数为 λp 的泊松分布。

解： 用 Y 表示商场内一天购物的顾客数，则由全概率公式知，对任意正整数 k 有

$$P(Y = k) = \sum_{i=k}^{\infty} P(X = i) P(Y = k \mid X = i) = \sum_{i=k}^{\infty} \frac{\lambda^i e^{-\lambda}}{i!} C_i^k p^k (1-p)^{i-k}$$

$$= \frac{(\lambda p)^k}{k!} e^{-\lambda} \sum_{i=k}^{\infty} \frac{[\lambda(1-p)]^{i-k}}{(i-k)!} = \frac{(\lambda p)^k}{k!} e^{-\lambda} e^{\lambda(1-p)} = \frac{(\lambda p)^k}{k!} e^{-\lambda p}$$

□

3.2.4 超几何分布

定义： 若随机变量 X 的概率分布满足:

$$P(X=k)=\frac{C_D^k C_{N-D}^{n-k}}{C_N^n},\quad k=0,1,\cdots,n.$$

则称 X 服从超几何分布，记为 $X\sim H(N,D,n)$ （参数 N,D,n 满足 $N\geqslant D\geqslant 0$）

设一批产品有 N 个产品, D 个次品, 任取 n 个, 抽到的次品数为 X。如果进行放回抽样则 X 服从二项分布，如果进行不放回抽样则 X 服从超几何分布。

定理 2.3： 给定 n. 当 $N\to\infty,\frac{D}{N}\to p$ 时,

$$\lim_{N\to\infty}\frac{C_D^k C_{N-D}^{n-k}}{C_N^n}=C_n^k p^k(1-p)^{n-k},\quad k\geqslant 0$$

证明：由于 $0<p<1$，当 N 充分大时，$n<D<N$，且 n 是固定的，易知

$$\begin{aligned}\frac{C_D^k C_{N-D}^{n-k}}{C_N^n}&=\frac{D!}{k!(D-k)!}\cdot\frac{(N-D)!}{(n-k)!(N-D-n+k)!}\cdot\frac{n!(N-n)!}{N!}\\&=\frac{n!}{k!(n-k)!}\cdot\frac{D(D-1)\cdots(D-k+1)}{N^k}\\&\quad\cdot\frac{(N-D)(N-D-1)\cdots(N-D-n+k+1)}{N^{n-k}}\\&\quad\cdot\frac{N^n}{N(N-1)\cdots(N-n+1)}\\&=C_n^k\left(\prod_{i=1}^{k}\frac{D-i+1}{N}\right)\left(\prod_{i=1}^{n-k}\frac{N-D-i+1}{N}\right)\left(\prod_{i=1}^{n}\frac{N}{N-i+1}\right)\\&\to C_n^k p^k(1-p)^{n-k}\quad(N\to\infty)\end{aligned}$$

□

该定理的直观解释是，如果一批产品的总量 N 很大，其中次品占比为 p，则从整批产品随机抽取 n 个，抽到次品的个数 k 近似服从参数为 p，n 的二项分布。

3.2.5 几何分布

定义： 若随机变量 X 的所有可能值是全体整数，且概率分布满足:

$$P(X=k)=(1-p)^{k-1}p,\quad k=1,2,\cdots.$$

则称 X 服从几何分布，记为 $X\sim G(p)$, 参数 $0<p<1$ 。

例如，某个射手向目标连续射击，如果他单次射中目标的概率为 p，则他首次射中目标所需要的射击次数 X 是一个随机变量，且满足几何分布。

几何分布具备无记忆性: $P(X-n=k \mid X>n)=P(X=k)$.

例： 设 X 是只取自然数的离散随机变量，若 X 的分布具有无记忆性，证明 X 的分布一定为几何分布。

证明： 由无记忆性知

$$P(X>n+m|X>m)=\frac{P(X>n+m)}{P(X>m)}=P(X>n),$$

将 n 换为 $n-1$ 仍有

$$P(X>n+m-1)=P(X>n-1)P(X>m).$$

两式相减有

$$P(X=n+m)=P(X=n)P(X>m).$$

设 $P(X=1)=p$，若取 $n=m=1$ 有

$$P(X=2)=p(1-p).$$

若取 $n=2$，$m=1$ 则有

$$P(X=3)=P(X=2)P(X>1)=p(1-p)^2.$$

若令 $P(X=k)=p(1-p)^{k-1}$，则用数学归纳法得

$$P(X=k+1)=P(X=k)P(X>1)=p(1-p)^k,\quad k=0,1,\cdots.$$

这表明 X 的分布为几何分布。 □

3.2.6 离散均匀分布

定义： 若随机变量 X 的概率分布满足:

$$P(X=k)=\frac{1}{N},\quad k=1,\cdots,N.$$

则称 X 服从离散均匀分布。

3.3 连续随机变量

定义： 连续型随机变量指: 存在 $p(x)$ 使得

$$P(a \leqslant X \leqslant b) = \int_a^b p(x)dx, \quad \forall a < b.$$

称 $p(\cdot)$ 为 X 的概率密度 (函数), 也记为 $p_X(\cdot)$.

连续随机变量有以下性质:

（1）非负: $p(x) \geqslant 0$

（2）规范: $\int_{-\infty}^{\infty} p(x)dx = 1$

（3）$P(X = x) = 0$ 在任意一点选中的概率都为 0.

（4）$p(\cdot)$ 在 x 连续, 即 $P(X \in [x, x + \Delta x]) = p(x)\Delta x + o(\Delta x)$,

以下是常见的连续随机变量:

3.3.1 均匀分布

定义： 如果随机变量 X 的分布密度为:

$$p(x) = \begin{cases} \frac{1}{b-a}; & 若 a \leqslant x \leqslant b; \\ 0, & 其他. \end{cases}$$

则称 X 服从区间 $[a, b]$（或 (a, b)）上的均匀分布，记为 $X \sim U(a, b)$ (参数 $a < b$) :

均匀分布的分布函数也可以写为 $p(x) = \frac{1}{b-a} 1_{\{a \leqslant x \leqslant b\}}$.

例如，某公共汽车站每隔 10 分钟会有一班公交车到达，一位搭乘该车的乘客在任意时刻到达车站是等可能的，则他的候车时间 X 是一个随机变量，且满足 $[0, 10]$ 上的均匀分布。

3.3.2 指数分布

定义： 如果随机变量 X 的分布密度为:

$$p(x) = \lambda e^{-\lambda x}, \quad x > 0.$$

则称 X 服从参数为 λ 的指数分布，记为 $X \sim \mathrm{Exp}(\lambda)$(参数 $\lambda > 0$)

若 X 服从参数为 λ 的指数分布，则对任何 $0 \leqslant a < b$ 有：

$$P(a < X < b) = \lambda \int_a^b e^{-\lambda x} dx = e^{-\lambda a} - e^{-\lambda b}$$

$$P(X > a) = e^{-\lambda a}$$

定理： (无记忆性): $P(X - s > t \mid X > s) = e^{-\lambda t}, \forall t, s \geqslant 0.$

不难看出，$P(X - s > t \mid X > s) = \frac{P(X-s>t)}{P(X>s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda t} = P(X > t)$

注意到，无记忆性是指数分布独有的，即设 X 是非负的随机变量，$P(X - s > t \mid X > s) = e^{-\lambda t}$ 对 $\forall t, s \geqslant 0$ 恒成立的充分必要条件是 X 服从指数分布。

证明：之前已经证明了充分性，现只需证明必要性：设 X 是非负随机变量满足 $P(X - s > t \mid X > s) = e^{-\lambda t}$，则

$$P(X > s) > 0, \quad P(X > s + t) = P(X > s)P(X > t)$$

令 $f(u) = P(X > u)$，则 $f(s + t) = f(s)f(t)$

于是 $f(1) = f(\frac{1}{n} \times n) = (f(\frac{1}{n}))^n$

从而 $f(\frac{m}{n}) = f(\frac{1}{n} \times m) = (f(\frac{1}{n}))^m = (f(1))^{\frac{m}{n}}$

故对任意正有理数 r，有 $f(r) = (f(1))^r$。由于 $0 < f(1) < 1$ 且 $f(u)$ 是关于 u 的减函数，因此对任意 $u \geqslant 0$，有 $f(u) = (f(1))^u$。

令 $\lambda = -\ln f(1)$，则 $f(u) = e^{-\lambda u}$，即

$$P(X > u) = e^{-\lambda u} = \int_{+\infty}^u e^{-\lambda x} dx$$

$$P(a < X < b) = \int_a^b \lambda e^{-\lambda x} dx \quad (0 \leqslant a < b)$$

说明 X 服从指数分布。 □

3.3.3 正态分布

定义： 如果随机变量 X 的分布密度为：

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\}$$

则称 X 服从参数为 μ，σ 的正态分布，记为 $X \sim N\left(\mu, \sigma^2\right)$ (参数 $\mu \in \mathbb{R}, \sigma > 0$)

参数 $\mu = 0$，$\sigma^2 = 1$ 时的正态分布称为标准正态分布 $N(0,1)$ ，分布密度是:

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

归一性： $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1$:

设 $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, 将积分的平方写为二重积分:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \times \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \frac{1}{2\pi} \iint_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2}} dxdy.$$

做极坐标变换:

$$x = r\cos\theta, y = r\sin\theta \Rightarrow \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} \end{vmatrix} = r.$$

因此，二重积分可以写为

$$\frac{1}{2\pi} \int_0^{2\pi} \left(\int_0^{\infty} e^{-\frac{r^2}{2}} r dr \right) d\theta = \int_0^{\infty} e^{-R} dR = 1$$

对于其他正态分布的密度函数 $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ ：令 $y = \frac{x-\mu}{\sigma}$, 则

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 1.$$

定义函数 Φ :

$$\Phi(x) = \int_{-\infty}^{x} \phi(x) dx.$$

容易看出 $\Phi(-x) = 1 - \Phi(x)$.

定理： 令 $x^* = \frac{x-\mu}{\sigma}$, 则

$$P(a < X < b) = \int_a^b \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right) dx = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

推论： 设随机变量 $X \sim N(\mu, \sigma^2)$，则对一切正数 k，有

$$P(\mu - k\sigma < X < \mu + k\sigma) = \Phi(k) - \Phi(-k) = 2\Phi(k) - 1$$

例如查表得 $\Phi(3) = 0.9987$, 因此

$$P(\mu-3\sigma<X<\mu+3\sigma)=\Phi(3)-\Phi(-3)=0.9974$$

该结果说明正态随机变量 X 的取值基本落在区间 $(\mu-3\sigma,\mu+3\sigma)$ 内。

3.3.4 伽马分布

定义： 如果随机变量 X 的分布密度为:

$$p(x)=\frac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x},\quad x>0.$$

则称随机变量 X 服从伽马分布，记为 $X\sim\Gamma(\alpha,\beta)$(参数 $\alpha,\beta>0$)

其中，称 $\Gamma(\alpha)=\int_0^{\infty}y^{\alpha-1}e^{-y}dy$ 为 Γ 函数。

若 $\Gamma(\alpha)$ 为 Γ 函数，则函数具备以下性质:

（1）$\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$

证明:

$$\int_0^{\infty}y^{\alpha}e^{-y}dy=-\left.y^{\alpha}e^{-y}\right|_0^{\infty}+\int_0^{\infty}\alpha y^{\alpha-1}e^{-y}dy=\alpha\int_0^{\infty}y^{\alpha-1}e^{-y}dy$$

□

（2）$\Gamma(1)=1;\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$

证明:

$$\Gamma\left(\frac{1}{2}\right)=\int_0^{\infty}\frac{1}{\sqrt{y}}e^{-y}dy=\sqrt{2}\int_0^{\infty}e^{-\frac{x^2}{2}}dx=\sqrt{\pi}.$$

□

（3）$\alpha=1$ 时就是指数分布参数为 β.

3.4 随机变量的严格定义

定义： 假设 $(\Omega,\mathscr{F},P)$ 是概率空间, $X:\Omega\to\mathbb{R}$ 满足:

$$\text{对任意 } x\in\mathbb{R} \text{ 都有 } \{X\leqslant x\}\in\mathscr{F},$$

则称 X 是一个随机变量.

定义： 令 $F(x)=P(X\leqslant x),x\in\mathbb{R}$. 称 F 为随机变量 X 的分布函数, 也记为 F_X.

定理: 分布函数 $F = F_X$ 的三条性质:

(1) 单调性: 若 $x \leqslant y$, 则 $F(x) \leqslant F(y)$.

(2) 规范性: $\lim_{x\to-\infty} F(x) = 0; \lim_{x\to\infty} F(x) = 1$.

(3) 右连续性: $\lim_{y\to x+} F(y) = F(x)$.

- 离散型: $P(X = x_i) = p_i$.　　x_i 为 F_X 的跳点, p_i 为跳跃幅度.
- 连续型: $F_X(x) = \int_{-\infty}^{x} p(z)dz$, 且

$$p(x) = F_X'(x).$$

反过来, 若 F_X “几乎”连续可导, 则为连续型 (定理 4.3, 4.4).

- 尾分布函数: $G(x) = P(X > x) = 1 - F(x)$.

  连续型: $p(x) = -G'(x)$.
- 例. $X \sim \mathrm{Exp}(\lambda)$.

$$G(x) = e^{-\lambda x}, \quad \forall x > 0,$$
$$\Rightarrow G'(x) = -\lambda G(x). \quad \lambda: \text{速率}.$$

- 由 $F_X(x)$ 可求出 $P(X \in B), \forall B$.
- 若 $F_X = F_Y$, 则称 X 与 Y 同分布, 记为 $X \overset{d}{=} Y$.
- $X = Y$, 即 $P(X = Y) = 1$, 则 $F_X = F_Y$. 反之不然.