

AI 中的数学

第二讲

方聪，概率统计部分参考章复熹和张原老师课件

2023 年秋季

- ① 条件概率
- ② 全概率与贝叶斯公式
- ③ 独立试验序列

- ① 条件概率
- ② 全概率与贝叶斯公式
- ③ 独立试验序列

条件概率

例题 5.1: 3 个白球和 2 个红球, 随机取两个, 已知第一个是白球, 问第二个也是白球的概率?

条件概率

例题 5.1: 3 个白球和 2 个红球, 随机取两个, 已知第一个是白球, 问第二个也是白球的概率?

	(1, 2)	(1, 3)	(1, 4)	(1, 5)
(2, 1)		(2, 3)	(2, 4)	(2, 5)
(3, 1)	(3, 2)		(3, 4)	(3, 5)
(4, 1)	(4, 2)	(4, 3)		(4, 5)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	

$P(A)$ 前 3 行, $P(AB)$ 前 3 行 3 列

$$P(B | A) = \frac{6}{12} = 0.5$$

条件概率

如果 $P(A) > 0$, $P(B | A) = \frac{P(AB)}{P(A)}$ 为已知 A 条件下, B 发生的条件概率

- 权重重新分配 $P(B) = P(B | \Omega)$
- 图解释
- 乘法公式: $P(AB) = P(B)P(A | B)$

乘法公式

- $P(B | A) = \frac{P(AB)}{P(A)} \quad (P(A) > 0)$
- $P(AB) = P(B)P(A | B)$
- $P(A_1 \dots A_n) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 A_2) \dots P(A_n | A_1 \dots A_{n-1})$

例 5.4 将 52 张牌随机均分 4 堆, 求: 各堆都含 Ace 的概率.

例 5.4 将 52 张牌随机均分 4 堆, 求: 各堆都含 Ace 的概率.

- 解法一、分别将桃杏梅方 Ace 称为 A-1, A-2, A-3, A-4. 记 E_{k,i_k} = 第 k 堆含 A- i_k 但不含其它 Ace.
- $E_{i_1 i_2 i_3 i_4} = E_{1,i_1} E_{2,i_2} E_{3,i_3} E_{4,i_4}$,

$$P(E_{i_1 i_2 i_3 i_4}) = \frac{C_{48}^{12}}{C_{52}^{13}} \cdot \frac{C_{36}^{12}}{C_{39}^{13}} \cdot \frac{C_{24}^{12}}{C_{26}^{13}}, \quad i_1, i_2, i_3, i_4 \text{ 互不相等}$$

- 当 (i_1, i_2, i_3, i_4) 取遍 1, 2, 3, 4 的全排时, 所有 $E_{i_1 i_2 i_3 i_4}$ 互不相交, 并起来为 E . 因此,

$$P(E) = 4! \times \star\star = \frac{3 \times 2 \times 13^3}{51 \times 50 \times 49}.$$

条件概率

(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, 则三种概率分别为多少?

条件概率

(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, 则三种概率分别为多少?

$$\begin{aligned} P(B_1 R_2 R_3) &= P(B_1) P(R_2 | B_1) P(R_3 | B_1 R_2) \\ &= \frac{b}{b+r} \cdot \frac{r+d}{b+r+c+d} \cdot \frac{r+d+c}{b+r+2c+2d}, \end{aligned}$$

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

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

以上概率与黑球在第几次被抽出有关。

条件概率

(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

$$P(B_1 R_2 R_3) = \cdot = \cdot = \frac{br(r-1)}{(b+r)(b+r-1)(b+r-2)}$$

(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

$$P(B_1 R_2 R_3) = \cdot = \cdot = \frac{br^2}{(b+r)^3}$$

条件概率

(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

$$P(B_1 R_2 R_3) = \cdot = \cdot = \frac{br(r+c)}{(b+r)(b+r+c)(b+r+2c)}$$

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, 下一次发生事故的的概率会增大, 上述三种概率分别为:

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

条件概率

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。

条件概率

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

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

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

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

$$P(AB) = P(\text{取出两件都是合格品}) = \frac{C_{n-m}^2}{C_n^2} = \frac{(n-m)(n-m-1)}{n(n-1)}$$

于是所求概率为

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

□

独立性

- 定义 5.2: 若 $P(AB) = P(A)P(B)$, 则称 A, B 相互独立
- 含义: A (B) 发生与否与 B (A) 无关, A (B) 不能影响 B (A) 的概率
- $P(B | A) = P(B)$
- 举例: 5 个乒乓球 (3 新, 2 旧) 有放回地取两次, $A =$ 第一次取的新球, $B =$ 第二次取的新球
- 投硬币
- 数据同分布抽样
- 若改为不放回地取两次呢?

独立性

- 定义 5.2: 若 $P(AB) = P(A)P(B)$, 则称 A, B 相互独立
- 含义: A (B) 发生与否与 B (A) 无关, A (B) 不能影响 B (A) 的概率
- $P(B | A) = P(B)$
- 举例: 5 个乒乓球 (3 新, 2 旧) 有放回地取两次, $A =$ 第一次取的新球, $B =$ 第二次取的新球
- 投硬币
- 数据同分布抽样
- 若改为不放回地取两次呢?
- 则改为 $P(B | A) = \frac{2}{4} \neq P(B)$.

因此, $P(AB) \neq P(A)P(B)$, 即 A 与 B 不独立

独立性

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

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

所以

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

故

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

独立性

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。

(2) 不放回情形。易知

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

故

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

独立性

- 若 A 与 B 独立, 那么 A 与 B^c 独立 (定理 5.2)

独立性

- 若 A 与 B 独立, 那么 A 与 B^c 独立 (定理 5.2)

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

多事件相互独立:

- 定义: 事件 A_1, \dots, A_n ($n \geq 2$) 是相互独立的, 如果对于任何整数 k ($2 \leq k \leq n$), 有

$$P(A_{i_1} \cdots A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k})$$

其中 i_1, \dots, i_k 是 $1 \leq i_1 < \dots < i_k \leq n$ 的任何 k 个整数

独立性

注意：独立 \Rightarrow 两两独立，但是反之不对：

伯恩斯反例：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

例 5.9: 每门炮击中敌机的概率是 0.6, 则要以 99% 的把握击中敌机, 需要至少多少高射炮?

独立性

例 5.9: 每门炮其中敌机的概率是 0.6, 则要以 99% 的把握击中敌机, 需要至少多少高射炮?

- 一门没有击中概率是 0.4
- n 门炮弹没有击中概率: $1 - 0.4^n$
- n 不是随机数
- $1 - 0.4^n \geq 0.99$, n 为整数
- $n \geq \ln(0.01)/\ln(0.4)$, 解为 $n = 6$
- 失败的概率指数减小

独立性

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。

□

- ① 条件概率
- ② 全概率与贝叶斯公式
- ③ 独立试验序列

全概率公式

- 假设 B_1, \dots, B_n 是 Ω 的分割, 则

$$P(A) = \sum_{i=1}^n P(B_i)P(A | B_i)$$

- B_1, \dots, B_n 是完备事件组
- 可改成可列分割

全概率公式

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？

全概率公式

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, \bar{B} 构成完备事件组，有

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

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

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

例 6.1. 甲、乙、丙三人射击敌机. 甲、乙、丙击中概率分别为: 0.4, 0.5, 0.7. 恰有 0, 1, 2, 3 人击中时飞机坠毁概率分别为 0, 0.2, 0.6, 1. 求: 飞机坠毁概率

例 6.1. 甲、乙、丙三人射击敌机. 甲、乙、丙击中概率分别为: 0.4, 0.5, 0.7. 恰有 0, 1, 2, 3 人击中时飞机坠毁概率分别为 0, 0.2, 0.6, 1. 求: 飞机坠毁概率

- $A =$ “飞机坠毁”, $B_i =$ “恰好 i 人击中”
- $P(B_0) = (1 - 0.4)(1 - 0.5)(1 - 0.7) = 0.09$,
 $P(B_1) = 0.4(1 - 0.5)(1 - 0.7) + (1 - 0.4)0.5(1 - 0.7)$

$$+ (1 - 0.4)(1 - 0.5)0.7 = 0.36,$$

$$P(B_2) = \dots = 0.41,$$

$$P(B_3) = 0.4 \cdot 0.5 \cdot 0.7 = 0.14.$$

- $P(A | B_i)$ 依次为 0, 0.2, 0.6, 1
- $P(A) = \sum_{i=0}^3 P(B_i) P(A | B_i)$

$$= 0.09 \times 0 + 0.36 \times 0.2 + 0.41 \times 0.6 + 0.14 \times 1 = 0.458$$

全概率公式

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。

全概率公式

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 $p_i = P(A_i), i = 0, 1, 2, \dots$ ，则有 $p_0 = 1$ ，且

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

由全概率公式得

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

得到递推公式

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

全概率公式

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

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



贝叶斯公式

- 假设 B_1, \dots, B_n 是 Ω 的分割, 则

$$P(B_k | A) = \frac{P(B_k)P(A | B_k)}{\sum_{i=1}^n P(B_i)P(A | B_i)}$$

- 条件概率, 乘法公式, 全概率公式
- 先验概率; 后验概率

例 6.8. 某艾滋病检测法的灵敏度如下:

假阴性的概率为 0.05; 假阳性的概率为 0.01.

已知某地区有 0.001 比例人群被艾滋病感染, 检测出甲被感染.

求: 甲感染艾滋病的概率.

- 显明: $T =$ 甲检测出被感染, $T^c =$ 甲检测出健康;
- 隐藏: $A =$ 甲被感染, $A^c =$ 甲健康.
- 已知: $P(A) = 0.001 = p$, $P(A^c) = 1 - p$.

$$P(T | A) = 1 - 0.05, P(T | A^c) = 0.01.$$

- 逆概公式:

$$\begin{aligned} P(A | T) &= \frac{P(A)P(T | A)}{P(A)P(T | A) + P(A^c)P(T | A^c)} \\ &= \frac{0.95p}{0.95p + 0.01(1 - p)} = \frac{0.95p}{0.94p + 0.01} \approx 0.087. \end{aligned}$$

- ① 条件概率
- ② 全概率与贝叶斯公式
- ③ 独立试验序列

独立试验序列

- 概率模型
- 每次实验 A 发生概率是 p , n 次独立重复试验事件 A 发生 k 次概率
- 例题 7.1

例 7.4. 甲、乙两人比赛, 每局甲赢的概率为 p , 乙赢的概率为 $q = 1 - p$, 赢者得 1 分, 输者得 0 分. 累计多 2 分者胜出. 求: 甲胜出的概率

例 7.4. 甲、乙两人比赛, 每局甲赢的概率为 p , 乙赢的概率为 $q = 1 - p$, 赢者得 1 分, 输者得 0 分. 累计多 2 分者胜出. 求: 甲胜出的概率

- $A =$ “甲胜出”, $B =$ “头两局甲赢”, $\tilde{B} =$ “头两局乙赢”,
 $C =$ “头两局甲、乙各赢一局”。
- $A = B \cup (CA)$, 于是

$$P(A) = P(B) + P(C)P(A | C) = P(B) + P(C)P(A)$$

- $P(B) = p^2, P(C) = 2pq$
- 解得:

$$P(A) = \frac{p^2}{1 - 2pq} = \frac{p^2}{p^2 + q^2} = P(B | B \cup \tilde{B})$$