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

2024 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？

$$\begin{array}{ccccc}
 & (1,2) & (1,3) & (1,4) & (1,5) \\
(2,1) & & (2,3) & (2,4) & (2,5) \\
(3,1) & (3,2) & & (3,4) & (3,5) \\
(4,1) & (4,2) & (4,3) & & (4,5) \\
(5,1) & (5,2) & (5,3) & (5,4) &
\end{array}$$

$P(A)$ 前 3 行，$P(AB)$ 前 3 行 3 列

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

条件概率

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

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

乘法公式

- $P(B \mid A) = \frac{P(AB)}{P(A)}\ (P(A) > 0)$
- $P(AB) = P(B)P(A \mid B)$
- $P(A_1 \dots A_n) = P(A_1)P(A_2 \mid A_1)P(A_3|A_1A_2) \cdots P(A_n \mid A_1 \cdots A_{n-1})$

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 $\mathrm{A}-i_k$ 但不含其它 Ace.
- $E_{i_1 i_2 i_3 i_4} = E_{1,i_1} E_{2,i_2} E_{3,i_3} E_{4,i_4}$,

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

- 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？

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

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

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

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

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

（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

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

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

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

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：

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

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

$P(A)=P$（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

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

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
- $P(B \mid A) = P(B)$
- 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
- $P(B \mid A) = P(B)$
- 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 $P(B \mid A) = \frac{2}{4} \neq P(B)$.

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

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

所以

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

故

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

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

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

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 $A_1, \cdots, A_n$ $(n \geq 2)$ 是相互独立的，如果对于任何整数 $k(2 \geq k \geq n)$，有

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

其中 $i_1, \cdots i_k$ 是 $1 \geq i_1 < \cdots < i_k \geq n$ 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

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？

- 一门没有击中概率是 0.4
- n 门炮弹没有击中概率：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。 $\square$

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 $B_1, \cdots, B_n$ 是 Ω 的分割，则

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

- $B_1, \cdots, B_n$ 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 $B,\overline{B}$ 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

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

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"
- $P(B_0)=(1-0.4)(1-0.5)(1-0.7)=0.09$,
  $P(B_1)=0.4(1-0.5)(1-0.7)+(1-0.4)0.5(1-0.7)$
  $$+(1-0.4)(1-0.5)0.7=0.36,$$
  $P(B_2)=\cdots=0.41$,
  $P(B_3)=0.4\cdot 0.5\cdot 0.7=0.14$.
- $P(A\mid B_i)$ 依次为 $0, 0.2, 0.6, 1$
- $P(A)=\sum_{i=0}^{3} P(B_i)P(A\mid B_i)$

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

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 $p_i = P(A), i = 0, 1, 2, \cdots$，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

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

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

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

$\square$

贝叶斯公式

- 假设 $B_1, \cdots B_n$ 是 Ω 的分割，则

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

- 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.

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

$$P(T \mid A)=1-0.05, P\left(T \mid A^c\right)=0.01.$$

- 逆概公式:

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

贝叶斯公式

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。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？

贝叶斯公式

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。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

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

$$= \frac{0.95\times0.005}{0.95\times0.005 + 0.01\times0.995}$$

$$= \frac{95}{294} \approx 0.323$$

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 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

$$P(A)=P(B)+P(C)P(A\mid C)=P(B)+P(C)P(A)$$

- $P(B)=p^2, P(C)=2pq$
- 解得:

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

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：

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：

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

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

随机游走

为使 x 为整数，k 和 n 的奇偶性需要相同，即

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

随机游走

为使 x 为整数，k 和 n 的奇偶性需要相同，即

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

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：

随机游走

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

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

随机游走

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

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

随机游走

(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

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

结合边值条件，得

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

则

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