第二章、条件概率与统计独立性

§2.1 条件概率, 全概率公式, 贝叶斯公式

例. 有10 个球, 5 个玻璃球(2 黑, 3 红), 5 个木球(4 黑, 1 红). 从中任取一个. 记$A =$ “取到玻璃球”, $B =$ “取到红球”.

- $P(B) = \frac{4}{10} = \frac{2}{5}$.
- 已知取到玻璃球/木球, 求: 取到红球的概率.
- $P(B|A) = \frac{3}{5} > P(B)$,

  $P(B|A^c) = \frac{1}{5} < P(B)$.

	玻璃	木质	
黑	2	4	6
红	3	1	4
	5	5	10

- 假设$(\Omega, \mathcal{F}, P)$ 是一个概率空间, $B \in \mathcal{F}$ 且$P(B) > 0$.

  $\forall A \in \mathcal{F}$, 称

$$\frac{P(AB)}{P(B)}$$

  为在B 发生的条件下, A 的条件概率,

  记为$P(A|B)$ 或$P_B(A)$. (定义2.1.1)

- $P_B(\cdot) = P(\cdot|B)$ 满足概率定义的三个条件.

- 应用一、按照定义直接计算$P(A|B)=P(AB)/P(B)$.
- 若P 是古典概型, 则P_B 是古典概型:

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

- 已知B, 如果B, 假设B, 若B, 当B 时, 在B 中, $\cdots$.
- $P_B(\cdot|C)=P(A|BC)$,

$$P_B(A|C)=\frac{P_B(AC)}{P_B(C)}=\frac{P(ABC)/P(B)}{P(BC)/P(B)}=P(A|BC).$$

- 应用二、乘法公式.

  分析P_B: 在假设B 发生时, 简化模型, 获得$P(A|B)$.

$$P(AB) = P(B)P(A|B). \quad (2.1.2)$$

- n 个事件的乘法公式:

$$\begin{aligned} P(A_1\cdots A_n) =& P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots \\ & \times P(A_n|A_1\cdots A_{n-1}). \quad (2.1.4) \end{aligned}$$

例2.1.2. 波利亚坛子(Polya Urn):

最初有b 个黑球, r 个红球. 每次取一个, 放回并放入c 个同色球.

$$B_n=\text{“第}n\text{ 次抽到黑球”},\quad R_n=\text{“第}n\text{ 次抽到红球”}=B_n^c.$$

求: $P(B_1B_2R_3R_4)$, $P(B_n)$.

- $BBRR:=B_1B_2R_3R_4$, $BRB:=B_1R_2B_3$.
- $B_1R_3B_5=B*R*B\neq BRB$, $R_3R_5=**R*R$.
- 例,

$$P(BBRR)=\frac{b}{b+r}\cdot\frac{b+c}{b+r+c}\cdot\frac{r}{b+r+2c}\cdot\frac{r+c}{b+r+3c}.$$

- 又例,

$$P(BRBR)=\frac{b}{b+r}\cdot\frac{r}{b+r+c}\cdot\frac{b+c}{b+r+2c}\cdot\frac{r+c}{b+r+3c}.$$

- $BBRR$, $BRBR$, $RBBR$, $BRRB$, $RBRB$, $RRBB$ 的概率都相等. 可交换.
- 例, $B_3 = *\star B = BBB \cup BRB \cup RBB \cup RRB$.
- 可交换:

$$P(*\star B) = P(B*\star).$$

- 例, $P(B_3) = \sum_{*,\star} P(B*\star) = P(B)$:

$$\begin{aligned} P(B_3) =& P(BBB) + P(BRB) + P(RBB) + P(RRB) \\ =& P(BBB) + P(BBR) + P(BRB) + P(BRR) \end{aligned}$$

- $P(B_n) = P(B_1) = \frac{b}{b+r}$.

例. 将52张牌随机均分4组, 令$A=$“各组都含K”, 求$P(A)$.

- 解法一、 $A=A_1A_2A_3A_4$, 其中$A_i=$“第i 组有K”.

$$P(A)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)P(A_4|A_1A_2A_3)=?$$

- 解法二、 $A=B_1B_2B_3$, 其中$B_i=$“第i 组恰有一个K”.

$$P(A)=\frac{C_4^1C_{48}^{12}}{C_{52}^{13}}\cdot\frac{C_3^1C_{36}^{12}}{C_{39}^{13}}\cdot\frac{C_2^1C_{24}^{12}}{C_{26}^{13}}.$$

- 解法三、$A = C_1C_2C_3$.

(1) C_1 = 红心K 与黑桃K 不在一组;

(2) C_2 = 梅花K 与黑桃K、红心K 在三个不同组中;

(3) C_3 = 4 张K 在四个不同组中.

- 乘法公式:

$$\begin{aligned}P(A) &= P(C_1)P(C_2|C_1)P(C_3|C_2)\\ &= \frac{C_{50}^{12}}{C_{51}^{12}}\cdot\frac{C_{49}^{24}}{C_{50}^{24}}\cdot\frac{C_{48}^{36}}{C_{49}^{36}}\\ &= \frac{39}{51}\times\frac{26}{50}\times\frac{13}{49}.\end{aligned}$$

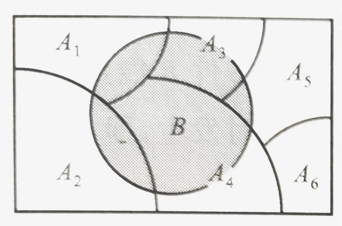
全概公式

- 全概公式: 假设$A_i,\ i \in I$ 是Ω 的一个可数划分, 则

$$P(B) = \sum_i P(A_i)P(B|A_i).$$

- $B = \sum_i (BA_i)$.
- $P(B) = \sum_i P(BA_i)$

  $= \sum_i P(A_i)P(B|A_i)$.



- 划分可改为:

$$P(A_iA_j) = 0, \forall i \neq j; \quad P\left(\bigcup_i A_i\right) = 1; \quad P(A_i) > 0, \forall i.$$

例2.1.2. 波利亚坛子(Polya Urn): 最初有b 个黑球, r 个红球. 每次取一个, 放回并放入c 个同色球. 求: $P(B_n)$.

- 用数学归纳法证明: $P(B_n) = P(B_1) \overset{\checkmark}{=} \frac{b}{b+r}$.
- 假设$P(B_{n-1}) = \frac{b}{b+r}$ 对任意正整数b 与r 成立. 那么,

$$\begin{aligned} P(B_n) =& P(B_1)P(B_n|B_1) + P(B_1^c)P(B_n|B_1^c) \\ =& \frac{b}{b+r} \cdot \frac{b+c}{(b+c)+r} + \frac{r}{b+r} \cdot \frac{b}{b+(r+c)} = \frac{b}{b+r}. \end{aligned}$$

- 上面方法称为“**首步分析法**”或“**向前分析法**”. 即, 根据第一步试验结果将Ω 划分为:

$$\Omega = B_1 + B_1^c.$$

P_{B_1} 与$P_{B_1^c}$ 分别变为参数为$(b+c, r)$ 与$(b, r+c)$ 的模型; 再利用全概公式.

例. 现有n 个球, n_1个红, n_2个黑. 从中任取m 个, 再从这m 个中任取r 个, 求: 这r 个中恰有r_1 个红球, r_2 个黑球的概率.

- 令A 表示r 个中恰有r_1 个红球, r_2 个黑球.
- B_{m_1} = 这m 个球中恰有m_1 个红球, $m_2 = m - m_1$ 个黑球.

$$P(B_{m_1}) = C_{n_1}^{m_1} C_{n_2}^{m_2} / C_n^m, \quad P(A|B_{m_1}) = C_{m_1}^{r_1} C_{m_2}^{r_2} / C_m^r.$$

- 乘法公式:

$$P(AB_{m_1}) = \frac{C_{n_1}^{m_1} C_{n_2}^{m_2} C_{m_1}^{r_1} C_{m_2}^{r_2}}{C_n^m C_m^r}.$$

-

$$P(A) = \sum_{m_1} P(AB_{m_1}) = \sum_{m_1} \frac{C_{n_1}^{m_1} C_{n_2}^{m_2} C_{m_1}^{r_1} C_{m_2}^{r_2}}{C_n^m C_m^r}. \quad \text{化简!}$$

- $P(A)=\sum_{m_1} C_{n_1}^{m_1} C_{m_1}^{r_1} C_{n_2}^{m_2} C_{m_2}^{r_2} / (C_n^m C_m^r)$.
- $C_n^m C_m^r = C_n^r C_{n-r}^{n-m}$:

$$
\begin{array}{ccccccccc}
 & & & \nearrow & r & \leftarrow & r & \leftarrow & n \\
n & \rightarrow & m & \rightarrow & m-r & \leftarrow & n-r & \swarrow & \\
 & \searrow & n-m & \rightarrow & n-m & \swarrow & & &
\end{array}
$$

一次性分成三份: $n-m,\ m-r,\ r$.

- 分子:

$$\sum_{m_1} C_{n_1}^{r_1} C_{n_1-r_1}^{n_1-m_1} C_{n_2}^{r_2} C_{n_2-r_2}^{n_2-m_2} = C_{n_1}^{r_1} C_{n_2}^{r_2} \sum_{m_1} C_{n_1-r_1}^{n_1-m_1} C_{n_2-r_2}^{n_2-m_2}.$$

- 分母: $C_n^r C_{n-r}^{n-m}$

(公式: 对任意$r \leqslant \min\{m,n\}$, $\sum_{k=0}^r C_m^{r-k} C_n^k = C_{n+m}^r$), 故

$$P(A)=\sum_{m_1} P(AB_{m_1}) = \frac{C_{n_1}^{r_1} C_{n_2}^{r_2}}{C_n^r}.$$

例. 将52 张牌随机均分两组. 从第一组中取一张(记为甲), 发现它是K, 将之放入第二组. 再从第二组(共27 张牌)中取出一张(记为乙), 求A =“乙是K” 的概率.

- 解法一、 记B_i =“第二组原有的26张牌中有i 张K”.

$$P(B_i) = \frac{C_4^i C_{48}^{26-i}}{C_{52}^{26}}, \quad P(A|B_i) = \frac{i+1}{27}.$$

故, 所求为 $P(A) = \sum_{i=0}^{3} \frac{C_4^i C_{48}^{26-i}}{C_{52}^{26}} \cdot \frac{i+1}{27}$. (错误)

- C =“甲是K”, 所求为

$$P_C(A) = P(A|C).$$

解法一(修正)、$A=$“乙是K”, $C=$“甲是K”.

- 全概公式:

$$P(C)=\sum_{i=0}^{3}P(B_i)P(C|B_i),$$

$$P(B_i)=\frac{C_4^iC_{48}^{26-i}}{C_{52}^{26}},\quad P(C|B_i)=\frac{4-i}{26}.$$

- 全概公式:

$$P(AC)=\sum_{i=0}^{3}P(B_i)P(AC|B_i)$$

$$*=\frac{4-i}{26}\star,\quad P(A|B_iC)=\frac{i+1}{27}.$$

- 条件概率的定义: $P(A|C)=P(AC)/P(C)=\cdots$.

解法二、$A=$“乙是K”, $C=$“甲是K”.

- 直接分三组: 26, 25, 1.

$$P_C(B_i)=C_3^iC_{48}^{26-i}/C_{51}^{26},\quad P_C(A|B_i)=P_{CB_i}(A)=\frac{i+1}{27},$$

$$P_C(A)=\sum_{i=0}^{3}P_C(B_i)P_C(A|B_i)=\sum_{i=0}^{3}\frac{C_3^iC_{48}^{26-i}}{C_{51}^{26}}\cdot\frac{i+1}{27}.$$

- 分子:

$$48!\left(\frac{1}{26!22!}+\frac{3\cdot 2}{25!23!}+\frac{3\cdot 3}{24!24!}+\frac{4}{23!25!}\right)=\cdots=\frac{12642\cdot 48!}{24!26!}.$$

- 所求为:

$$\begin{aligned}P_C(A)&=\frac{12642\cdot 48!}{24!26!}/\left(\frac{51!}{26!25!}\cdot 27\right)\\&=\frac{12642\cdot 25}{51\cdot 50\cdot 49\cdot 27}=\frac{258}{51\cdot 2\cdot 27}=\frac{43}{17\cdot 27}.\end{aligned}$$

解法三、$A=$“乙是K”, $C=$“甲是K”.

- 先任取出甲(1 张), 再将余牌(51 张)分两组: 25, 26.
- 在$C=$ “甲是K” 的条件下, 新模型: 将51 张牌(仅3 张K)随机均分两组: 25, 26. 从甲与第二组(共27 张牌)中取出乙.
- $B=$“乙是甲”, $B^c=$ “乙出自第二组” :

$$P_C(B)=\frac{1}{27},\quad P_C(A|B)=1;$$
$$P_C(B^c)=\frac{26}{27},\quad P_C(A|B^c)=?$$

- 根据对称性, 在$P_C(\cdot|B^c)$ 下, 乙从余牌中等可能随机取出.

$$P_C(A|B^c)=\frac{3}{51}=\frac{1}{17}.$$

- $P_C(A)=\frac{1}{27}+\frac{26}{27}\cdot\frac{1}{17}=\frac{17+26}{27\cdot 17}=\frac{43}{27\cdot 17}.$

贝叶斯公式

- 贝叶斯(Bayes)公式: 假设A_i, $i \in I$ 是Ω 的一个可数划分, 则

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_j P(A_j)P(B|A_j)}.$$

- 划分可改为:

$$P(A_iA_j) = 0, \forall i \neq j; \quad P\left(\bigcup_i A_i\right) = 1; \quad P(A_i) > 0, \forall i.$$

- 推导/计算: $P(A_i|B) = \frac{P(A_iB)}{P(B)}$ & 乘法公式& 全概公式.
- 逆概公式, 先验概率 vs 后验概率.
- B 是显明的, A_i 是隐藏的.

$$\mathcal{G} = \{\emptyset, B, B^c, \Omega\}. \ A_i \in \mathcal{F}, \text{但} A_i \notin \mathcal{G}.$$

例2.1.5. $C=$“有肝癌”, $A=$“被某肝癌检测法检测出阳性”.

$$P(A|C)=0.95;\ P(A^c|C^c)=0.9;\ P(C)=0.0004.$$

求: $P(C|A)$.

- $P(AC)=P(C)P(A|C)=0.0004\cdot 0.95=0.00038,$

  $P(AC^c)=P(C^c)P(A|C^c)=0.9996\cdot 0.1=0.09996.$

- 于是,

$$P(C|A)=\frac{P(AC)}{P(AC)+P(AC^c)}\approx 0.0038.$$

§2.2. 事件的独立性

- 直观: $P(B|A) = P(B)$, A 发生(与否)不改变B 的概率.
- 若

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

则称A, B (相互)独立(independent). (定义2.2.1)

- 设$0 < P(A) < 1$, $0 < P(B) < 1$, 则A, B 独立

$$\text{iff } P(A|B) = P(A) \quad \text{iff } P(A|B^c) = P(A),$$
$$\text{iff } P(B|A) = P(B) \quad \text{iff } P(B|A^c) = P(B).$$

- 若$P(A_iA_j) = P(A_i)P(A_j)$, $\forall i \neq j$, 则称A_i, $i \in I$ 两两独立.

例2.2.1 vs 例2.2.2. a 个黑球, b 个白球, 抽2 次.

$$A = \text{“第一次是黑”}, \quad B = \text{“第二次是黑”}.$$

- 样本空间:

$$\Omega = \{(i,j) : i, j \leqslant n\}, \quad n = a + b.$$

- 放回抽样:

$$p_\omega = \frac{1}{n^2}, \ \forall \omega, \quad P(AB) = \frac{a^2}{n^2} = P(A)P(B).$$

- 不放回抽样:

$$\tilde{p}_\omega = \frac{1}{n(n-1)}, \ \forall i \neq j, \quad \widetilde{P}(AB) = \frac{a(a-1)}{n(n-1)} < \widetilde{P}(A)\widetilde{P}(B).$$

习题一、5. “石头、剪刀、布” 游戏.

$A =$“甲出剪刀”, $B =$“乙出布”, $C =$“甲赢”.

- 样本: $\begin{matrix} (0,0) & (0,2)_C & (0,5)_B \\ (2,0)_A & (2,2)_A & (2,5)_{A,B,C} \\ (5,0)_C & (5,2) & (5,5)_B \end{matrix}$

- A, B, C 两两独立:

$$P(A) = P(B) = P(C) = \frac{1}{3},$$
$$P(AB) = P(AC) = P(BC) = \frac{1}{9}.$$

- 但, $P(C|AB) = 1$.

- 若$A_1, \cdots, A_n$ 满足:

$$P(A_{i_1} \cdots A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}),$$
$$\forall k \leqslant n, \quad \forall 1 \leqslant i_1 < \cdots < i_k \leqslant n,$$

则称它们相互独立. (定义2.2.3)

- 例. $n = 3$. (定义2.2.2)

$$P(ABC) = P(A)P(B)P(C) = P(AB)P(C).$$

- $A_1, A_2, \cdots$ 相互独立: $A_1, \cdots, A_n$ 相互独立, $\forall n$, iff

$$P(A_{i_1} \cdots A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}), \quad \forall 1 \leqslant i_1 < \cdots < i_k.$$

独立的性质： 若$A_1, \cdots, A_n$ 相互独立, 则

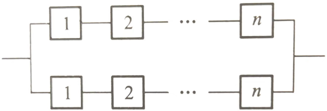
- $\forall 1 \leqslant j_1 < j_2 < \cdots < j_k \leqslant n$, $A_{j_1}, \cdots, A_{j_k}$ 相互独立;
- B_i 取A_i 或A_i^c, 则B_i, $1 \leqslant i \leqslant n$ 相互独立;
- $(A_1A_2), A_3, \cdots, A_n$ 相互独立;
- $(A_1 \cup A_2), A_3, \cdots, A_n$ 相互独立.

例2.2.6. $C_i=$“第i 个元件可靠”. $C_i,\ i\in I$ 相互独立, $P(C_i)\equiv r$.

求: P(系统可靠).

- $A_1=C_1\cdots C_n,\qquad A_2=C_{n+1}\cdots C_{2n}$.
- $R_1=P(A_1\cup A_2)$:

  A_1 与A_2 独立.



(a)

- 若当公式:

$$\begin{aligned}R_1=&P(A_1)+P(A_2)-P(A_1A_2)\\=&2r^n-r^{2n}=r^n(2-r^n).\end{aligned}$$

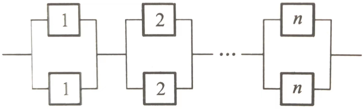
- 对偶公式:

$$\begin{aligned}1-R_1=&P(A_1^cA_2^c)=P(A_1^c)P(A_2^c)\\=&(1-r^n)^2=1-2r^n+r^{2n}.\end{aligned}$$

- $B_k = C_{2k-1} \cup C_{2k}$, $B_1, \cdots, B_n$ 相互独立.

- 若当公式:

  $P(B_k) = r + r - r^2 = r(2-r)$.



(b)

- $R_2 = P(B_1 \cdots B_n)$.

- $R_2 = P(B_1)^n = r^n(2-r)^n$.

- $R_2 > R_1$: $(2-r)^n > 2 - r^n$, $n \geqslant 2$.

§2.3 伯努利试验~~与直线上的随机游动~~

- (小)试验:

$$(\Omega_i, \mathcal{F}_i, P_i): \text{ 第}i\text{ 个小试验, } i = 1, \cdots, n(\text{或}i \geqslant 1).$$

- 大试验的样本空间Ω:

$$\Omega = \Omega_1 \times \cdots \times \Omega_n, \quad \omega = (\omega_1, \cdots, \omega_n).$$

- 事件: 设$\tilde{A}_i \in \mathcal{F}_i$,

$$\begin{aligned}\tilde{A}_i \hookrightarrow A_i =& \{\omega : \omega_i \in \tilde{A}_i\} \\ =& \Omega_1 \times \cdots \times \Omega_{i-1} \times \tilde{A}_i \times \Omega_{i+1} \times \cdots \times \Omega_n.\end{aligned}$$

- 取σ 代数$\mathcal{F} := \sigma(\{A_i : \tilde{A}_i \in \mathcal{F}_i, 1 \leqslant i \leqslant n\})$.

- $(\Omega, \mathcal{F})$ 存在唯一的概率P 满足:

(1) 与小试验相容:

$$P(A_i) = P_i(\tilde{A}_i), \quad \forall \tilde{A}_i \in \mathcal{F}_i,\ 1 \leqslant i \leqslant n.$$

(2) 小试验相互独立:

$$P(A_1 \cdots A_n) = \prod_i P(A_i), \quad \forall \tilde{A}_i \in \mathcal{F}_i,\ 1 \leqslant i \leqslant n.$$

例. 抽球两次.

- 两个小试验: 抽球一次,

$$\Omega_1 = \Omega_2 = \{1, \cdots, n\}, \quad P_1 = P_2 : \quad \{i\} \mapsto \frac{1}{n}.$$

- 大试验样本: $\omega = (i, j)$, $i, j \in \{1, \cdots, n\}$.
- 放回抽样: 小试验相互独立.

$$P : \{(i, j)\} \mapsto \frac{1}{n} \times \frac{1}{n}.$$

- 不放回抽样: 小试验不相互独立, $\tilde{P}(\cdot) = P(\cdot | i \neq j)$.

$$\tilde{P} : \{(i, j)\} \mapsto \frac{1}{n} \times \frac{1}{n-1}, \quad \forall j \neq i.$$

- $\hat{P} : \{(i, 1)\} \mapsto \frac{1}{n}, \quad \forall i.$ 与小试验不相容.

独立重复试验:

- 重复性: $(\Omega_i, \mathcal{F}_i, P_i) \equiv (\Omega_1, \mathcal{F}_1, P_1)$,
- 伯努利(Bernoulli)试验: $0 < p < 1$,

$$\forall i,\ \Omega_i = \{H, T\}, \quad P_i(\{H\}) = p, \quad P_i(\{T\}) = 1 - p =: q.$$

或者一般地, $\forall i$,

$$\Omega_i = \Omega_1, \quad \mathcal{F}_i = \{\varnothing, \hat{A}, \hat{A}^c, \Omega_i\}, \quad P_i(\hat{A}) = p.$$

- ω: 无穷长(或有限长)的H-T 字符串.

$$H_n = \text{“第}n\text{ 次投到}H\text{”}, \quad T_n = H_n^c.$$

- 取$\mathcal{F} = \sigma(\{H_n, n \geqslant 1\})$.
- 存在唯一的定义在$\mathcal{F}$上的概率P 使得:

$$P(H_n) = p \in (0,1) \quad \text{且} \quad H_1, H_2, \cdots \text{相互独立}.$$

- $A = \bigcup_{n=1}^{\infty} H_n$, $A^c = \bigcap_{n=1}^{\infty} T_n$. 则$P(A) = 1$:

$$P(A^c) \leqslant P\left(\bigcap_{n=1}^{N} T_n\right) = q^N \xrightarrow{N\to\infty} 0.$$

- (重复试验中)正概率事件一定(几乎必然)发生.