概率论与数理统计 (Fall 2024)
习题课讲义

2024 年 9 月 14 日

聂嵘琢, 祝国隽

B01GB006Y-02/Note2

---

1. 甲、乙两人参加某项目比赛, 甲胜的概率为 0.6 , 乙胜的概率为 0.4 . 有三种赛制供选择: 三局两胜、五局三胜和七局四胜. 问那种赛制对甲有利?

Ans: (1) 若采取三局两胜, 甲获胜后情形有: $A_1=\{$ 甲胜前 2 局 $\}$; $A_2=\{$ 前 2 局甲乙各胜 1 场，第 3 局甲胜 $\}$. 则 $P(A_1\cup A_2)=P(A_1)+P(A_2)=0.6^2+\binom{2}{1}\times 0.6\times 0.4\times 0.6=0.648$ 。

(2) 若采取五局三胜, 甲获胜后情形有: $B_1=\{$ 甲胜前 3 局 $\}$; $B_2=\{$ 前 3 局甲胜 2 场乙胜 1 场, 第 4 局甲胜 $\}$; $B_3=$ ” $\{$ 前 4 局甲胜 2 场乙胜 2 场, 第 5 局甲胜 $\}$. 则 $P(B_1\cup B_2\cup B_3)=P(B_1)+P(B_2)+P(B_3)=0.6^3+\binom{3}{2}\times 0.6^2\times 0.4\times 0.6+\binom{4}{2}\times 0.6^2\times 0.4^2\times 0.6=0.682$.

(3) 若采取七局四胜，甲获胜后情形有: $C_1=\{$ 甲胜前 4 局 $\}$; $C_2=\{$ 前 4 局甲胜 3 场乙胜 1 场，第 5 局甲胜 $\}$; $C_3=\{$ 前 5 局甲胜 3 场乙胜 2 场，第 6 局甲胜 $\}$; $C_4=\{$ 前 6 局甲胜 3 场乙胜 3 场，第 7 局甲胜 $\}$. 则

$$\begin{aligned}P(C_1\cup C_2\cup C_3\cup C_4)=&P(C_1)+P(C_2)+P(C_3)+P(C_4)\\=&0.6^4+\binom{4}{3}\times 0.6^3\times 0.4\times 0.6+\binom{5}{3}\times 0.6^3\times 0.4^2\times 0.6\\&+\binom{6}{3}\times 0.6^3\times 0.4^3\times 0.6\\=&0.710\end{aligned}$$

综上, 七局四胜对甲更有利.

2. 令 A,B,C 是三个事件, $A=\left(\frac{1}{8},\frac{5}{8}\right)$, $B=\left(\frac{3}{8},\frac{7}{8}\right)$, $C=\left(\frac{1}{8},\frac{3}{8}\right)\cup\left(\frac{5}{8},\frac{7}{8}\right)$, 且每个事件发生的概率等于其区间的长度. 求证: A,B,C 两两独立但不相互独立.

Ans: $P(A)=\frac{5}{8}-\frac{1}{8}=\frac{1}{2}$, $P(B)=\frac{7}{8}-\frac{3}{8}=\frac{1}{2}$, $P(C)=\left(\frac{3}{8}-\frac{1}{8}\right)-\left(\frac{7}{8}-\frac{5}{8}\right)=\frac{1}{2}$

由 $A\cap B=\left(\frac{3}{8},\frac{5}{8}\right)$, $A\cap C=\left(\frac{1}{8},\frac{3}{8}\right)$, $B\cap C=\left(\frac{3}{8},\frac{5}{8}\right)$, $A\cap B\cap C=\emptyset$. 则

$$\begin{aligned}
P(AB)&=\frac{1}{4}=P(A)P(B),\\
P(AC)&=\frac{1}{4}=P(A)P(C),\\
P(BC)&=\frac{1}{4}=P(B)P(C),\\
P(ABC)&=0\neq P(A)P(B)P(C).
\end{aligned}$$

故 ABC 两两独立但不相互独立.

3. 某商场的一批产品由甲、乙、丙三个厂家提供. 已知甲、乙、丙三厂的次品率分别为 1%, 2% 3%, 甲、乙、丙三厂分别给商场提供 60% 25% 、15% 的产品. 问从该商场购置此产品的废品率是多少?

Ans: 记 $A=\{$ 从商场购置的产品是废品 $\}$, $B_1=\{$ 产品由甲生成 $\}$, $B_2=\{$ 产品由乙生成 $\}$ $B_3=\{$ 产品由丙生成 $\}$. 由题意知,

$$P(B_1)=0.6, P(B_2)=0.25, P(B_3)=0.15$$
$$P(A\mid B_1)=0.01, P(A\mid B_2)=0.02, P(A\mid B_3)=0.03$$

由全概率公式，可得

$$P(A)=\sum_{i=1}^{3}P(A\mid B_i)P(B_i)=0.01\times 0.6+0.02\times 0.25+0.03\times 0.15=0.0155$$

4. 一盒子里有 n 个球, 其中有 n_1 个白球, n_2 个红球. 从中随机取出 n_1 个球并放进 n_1 个白球. 再从盒子里随机取出 n_1 个球, 求取的球都是白球的概率, $n=n_1+n_2, n_2 \geqslant n_1$.

Ans: 解: 记 $B_i=\{$ 第一次取出 i 个白球 $\}, i=0,1,..,n_1, A=\{$ 再从盒中取 n_1 个球, 都是白球 $\}$ 由题意

$$P\left(B_0\right)=\frac{\binom{n_2}{n_1}}{\binom{n}{n_1}}, \quad P\left(B_1\right)=\frac{\binom{n_1}{1}\binom{n_2}{n_1-1}}{\binom{n}{n_1}}, \quad \ldots \quad P\left(B_k\right)=\frac{\binom{n_1}{k}\binom{n_2}{n_1-k}}{\binom{n}{n_1}},$$

且

$$P\left(A \mid B_0\right)=\frac{\binom{2n_1}{n_1}}{\binom{n}{n_1}}, \quad P\left(A \mid B_1\right)=\frac{\binom{2n_1-1}{n_1}}{\binom{n}{n_1}}, \cdots, \quad P\left(A \mid B_k\right)=\frac{\binom{2n_1-k}{n_1}}{\binom{n}{n_1}}.$$

由全概率公式

$$P(A)=\sum_{k=0}^{n_1} P\left(A \mid B_k\right) P\left(B_k\right)=\sum_{k=0}^{n_1} \frac{\binom{n_1}{k}\binom{n_2}{n_1-k}\binom{2n_1-k}{n_1}}{\binom{n}{n_1}^2}$$

5. 某地区人群中吸烟者的占比是 0.3 . 已知吸烟者患肺癌的概率是 0.005 , 不吸烟者患肺癌的概率是 0.0005 . 该地区某人患肺癌, 求此患者是吸烟者的概率.

Ans: 记 $A=\{$ 患肺癌 $\}, B=\{$ 该人吸烟 $\}, \bar{B}=\{$ 该人不吸烟 $\}$. 有

$$P(B)=0.3, P(\bar{B})=0.7, P(A \mid B)=0.005, P(A \mid \bar{B})=0.0005$$

所求概率为

$$P(B \mid A)=\frac{P(A \mid B) P(B)}{P(A \mid B) P(B)+P(A \mid \bar{B}) P(\bar{B})}=\frac{0.3 \times 0.005}{0.3 \times 0.005+0.7 \times 0.0005} \approx 0.811$$

6. 某地区人群患某种疾病的概率为 0.002 . 当检查时, 由于医学条件等原因, 使得患者未必检测出阳性而健康者可能被检测出阳性的结果. 假定患者检测出阳性的概率是

0.98，而健康者检测出阳性的概率是 0.001．某人独立地检测了 4 次，发现 3 次阳性，1 次阴性。问此人是患者的概率是多少？

Ans: 令事件 $A=\{$ 检出阳性 $\}$，事件 $B=\{$ 某人是患者 $\}$，$C=\{$ 某人独立地检查了 4 次，发现 3 次阳性，1 次阴性 $\}$．则 $\bar{B}=\{$ 某人健康 $\}$。由题意知

$$P(B)=0.002, P(\bar{B})=0.998, P(A \mid B)=0.98, P(A \mid \bar{B})=0.001$$

根据 Bayes 公式，得

$$\begin{aligned}
P(B \mid C) &= \frac{P(C \mid B)P(B)}{P(C \mid B)P(B)+P(C \mid \bar{B})P(\bar{B})} \\
&= \frac{\binom{4}{3}0.98^3(1-0.98)\times 0.002}{\binom{4}{3}0.98^3(1-0.98)\times 0.002+\binom{4}{3}0.001^3(1-0.001)\times 0.998} \\
&= 0.9999735.
\end{aligned}$$

有些同学没有用 *Bayes* 公式，直接算 $P(B \mid C)=\frac{3}{4}P(B \mid A)+\frac{1}{4}P(B \mid \bar{A})$，想一想到底是哪错了？

以下是课外补充资料，参考李东风老师 (PKU) 的讲义

- **解**

  (1) 用 A_i 表示该盘在第 i 年没被打破, 则至今没被打破的概率是

$$\begin{aligned}p =&P(A_1A_2\cdots A_{500})\\ =&P(A_1)P(A_2|A_1)\cdots P(A_{500}|A_1A_2\cdots A_{499})\\ =&(1-0.03)^{500}\\ =&2.43\times 10^{-7},\end{aligned}$$

  被失手打破的概率是 $q=1-p=0.999999756$.

- (2)　用 B_j 表示第 j 件至今已被打破, $m=10000$, 则 $B_1,B_2,\cdots,B_m$ 相互独立.

- 这 1 万件至今都已经被失手打破的概率是

$$q_1 =P(\bigcap\nolimits_{j=1}^{m}B_j)=\prod_{j=1}^{m}P(B_j)=q^m=0.99757.$$

- 有这类青花麒麟盘流传至今的概率是 $p_1=1-q_1=0.00243$.

- 如果当时生产了五十万件, 则有这类青花麒麟盘流传至今的概率是 $p_{50}=0.1149$.

- 如果当时生产了五百万件, 则有这类青花麒麟盘流传至今的概率是 $p_{500}=0.7048$.

- 当然，这个模型里每年失手打破的概率都是 0.03 的假设过于粗糙，实际上随着现存总数的减少保护必然加强，打破概率变得很小。

1.6　全概率公式与 Bayes 公式

1.6.1　全概率公式

全概率公式

- **定理 6.1**(全概率公式)　如果事件 $A_1,A_2,\cdots,A_n$ 互不相容, $B\subset\bigcup_{j=1}^{n}A_j$, 则

$$P(B)=\sum_{j=1}^{n}P(A_j)P(B|A_j). \tag{6.1}$$

- **证明:** 因为 $B = B(\bigcup_{j=1}^{n} A_j) = \bigcup_{j=1}^{n}(BA_j)$, 用概率的有限可加性和乘法公式得到

$$
\begin{aligned}
P(B) =& P(B(\bigcup\nolimits_{j=1}^{n} A_j)) \\
=& P(\bigcup\nolimits_{j=1}^{n} (BA_j)) \\
=& \sum_{j=1}^{n} P(BA_j) \\
=& \sum_{j=1}^{n} P(A_j)P(B|A_j).
\end{aligned}
$$

全概率公式—完备事件组

- 全概率公式容易推广到可列个事件的情况 (见习题 1.14)).
- 如果事件 $A_1, A_2, \cdots, A_n$ 互不相容, $\bigcup_{j=1}^{n} A_j = \Omega$, 则称 $A_1, A_2, \cdots, A_n$ 是**完备事件组**, 这时 (6.1) 对任何事件 B 成立.
- A 和 $\overline{A}$ 总构成完备事件组, 所以总有

$$
P(B) = P(A)P(B|A) + P(\overline{A})P(B|\overline{A}). \tag{6.2}
$$

例 6.1(抽签问题)

- **例 6.1** (抽签问题) n 个签中有 m 个标有“中”, 证明无放回依次随机抽签时, 第 j 次抽中的概率是 m/n.
- **解** 用归纳法. 用 A_j 表示第 j 次抽中, 则对一切 m, n, 当 $m \le n$ 时, 有 $P(A_1) = m/n$.
- 设对一切 m, n, 当 $m \le n$ 时, 有 $P(A_{j-1}) = m/n$, 则有

$$
P(A_j|A_1) = \frac{m-1}{n-1}, \quad P(A_j|\overline{A}_1) = \frac{m}{n-1}.
$$

- 于是有

$$
\begin{aligned}
P(A_j) =& P(A_1)P(A_j|A_1) + P(\overline{A}_1)P(A_j|\overline{A}_1) \\
=& \frac{m}{n}\ \frac{m-1}{n-1}\ +\ \frac{n-m}{n}\ \frac{m}{n-1} \\
=& \frac{m}{n}, \quad 1 \le j \le n.
\end{aligned}
$$

例 6.2(敏感问题调查)

- **例 6.2** (敏感问题调查) 在调查家庭暴力 (或婚外恋、服用兴奋剂、吸毒等敏感问题) 所占家庭的比例 p 时, 被调查者往往不愿回答真相, 这使得调查数据失真.

- 为得到实际的 p 同时又不侵犯个人隐私, 调查人员将袋中放入比例是 p_0 的红球和比例是 $q_0 = 1 - p_0$ 的白球.

- 被调查者在袋中任取一球窥视后放回, 并承诺取得红球就讲真话, 取到白球就讲假话.

- 被调查者只需在匿名调查表中选“是” (有家庭暴力) 或“否”, 然后将表放入投票箱.

- 没人能知道被调查者是否讲真话和回答的是什么. 如果声称有家庭暴力的家庭比例是 p_1, 求真正有家庭暴力的比例 p.

- **解** 对任选的一个家庭, 用 B 表示回答“是”, 用 A 表示实际“是”. 利用全概率公式得到

$$\begin{aligned}
p_1 =& P(B) \quad (\text{回答“是”}) \\
=& P(B|A)P(A) + P(B|\overline{A})P(\overline{A}) \\
=& p_0 P(A) + q_0(1 - P(A)) \\
& \big(P(B|A)\ \text{即讲真话概率},\ P(B|\overline{A})\ \text{等于讲假话概率}\big) \\
=& p p_0 + q_0(1-p) = q_0 + (p_0 - q_0)p.
\end{aligned}$$

- 于是只要 $p_0 \neq q_0$, 则

$$p = P(A) = \frac{p_1 - q_0}{p_0 - q_0}.$$

- 实际问题中, p_1 是未知的, 需要经过调查得到. 假定调查了 n 个家庭, 其中有 k 个家庭回答“是”, 则可以用 $\hat{p}_1 = k/n$ 估计 p_1, 于是可以用

$$\hat{p} = \frac{\hat{p}_1 - q_0}{p_0 - q_0}$$

估计 p.

- 如果袋中装有 30 个红球，50 个白球，调查了 320 个家庭，其中有 195 个家庭回答“是”，则

$$\begin{aligned}&p_0 = 3/8, \quad q_0 = 5/8,\\&\hat{p}_1 = 195/320,\\&\hat{p} = \frac{195/320 - 5/8}{3/8 - 5/8} = 6.25\%.\end{aligned}$$

- 可以证明 $|p_0 - q_0|$ 越大, 得到的结论越可靠. 但是 $|p_0 - q_0|$ 太大时, 调查方案不易让被调查者接受.

例 6.3(赌徒破产模型)

- **例 6.3** (赌徒破产模型)　甲有本金 a 元, 决心再赢 b 元停止赌博. 设甲每局赢的概率是 $p = 1/2$, 每局输赢都是一元钱, 甲输光后停止赌博, 求甲输光的概率 $q(a)$.

- **解**　用 A 表示甲第一局赢, 用 B_k 表示甲有本金 k 元时最后输光.

- 由题意, $q(0) = 1$, $q(a+b) = 0$, 并且

$$\begin{aligned}q(k) =&P(B_k)\\=&P(A)P(B_k|A) + P(\overline{A})P(B_k|\overline{A})\\=&\frac{1}{2}P(B_{k+1}) + \frac{1}{2}P(B_{k-1})\\=&\frac{1}{2}q(k+1) + \frac{1}{2}q(k-1).\end{aligned}$$

- 于是有 $2q(k) = q(k+1) + q(k-1)$.

- 从而得到

$$q(k+1) - q(k) = q(k) - q(k-1) = \cdots = q(1) - q(0) = q(1) - 1.$$

- 上式两边对 $k = n-1, n-2, \cdots, 0$ 求和后得到,

$$q(n) - 1 = n(q(1) - 1). \tag{6.3}$$

- 取 $n = a + b$, 得到

$$0 - 1 = (a+b)(q(1) - 1), \quad q(1) - 1 = -1/(a+b).$$

- 最后由 (6.3) 得到:

$$q(a)=1+a(q(1)-1)=1-\frac{a}{a+b}=\frac{b}{b+a}. \tag{6.4}$$

- (6.4) 说明, 当甲的本金 a 有限, 则贪心 b 越大, 输光的概率越大, 如果一直赌下去 ($b\to\infty$), 必定输光.

1.6.2 Bayes 公式

Bayes 公式

- **定理 6.2**(Bayes 公式) 如果事件 $A_1,A_2,\cdots,A_n$ 互不相容, $B\subset\bigcup_{j=1}^n A_j$, 则 $P(B)>0$ 时, 有

$$P(A_j|B)=\frac{P(A_j)P(B|A_j)}{\sum_{i=1}^n P(A_i)P(B|A_i)},\quad 1\le j\le n. \tag{6.5}$$

- **证明** 由条件概率公式和全概率公式得到

$$P(A_j|B)=\frac{P(A_jB)}{P(B)}=\frac{P(A_j)P(B|A_j)}{\sum_{i=1}^n P(A_i)P(B|A_i)},\quad 1\le j\le n.$$

- 值得指出的是, 分子总是分母中的一项.
- 当 $A_1,A_2,\cdots,A_n$ 是完备事件组, $P(B)>0$ 时, (6.5) 成立.
- Bayes 公式也可以推广到可列个事件的情况 (见习题 1.21).
- 最常用到的 Bayes 公式是当 $P(B)>0$,

$$P(A|B)=\frac{P(A)P(B|A)}{P(A)P(B|A)+P(\overline{A})P(B|\overline{A})}. \tag{6.6}$$

例 6.4(疾病普查问题)

- **例 6.4** (疾病普查问题) 一种新方法对某种特定疾病的诊断准确率是 90%(有病被正确诊断和没病被正确诊断的概率都是 90%). 如果群体中这种病的发病率是 0.1% , 甲在身体普查中被诊断患病, 问甲的确患病的概率是多少?
- **解** 设 $A=$ 甲患病, $B=$ 甲被诊断有病.
- 根据题意, $P(A)=0.001$,

$$P(B|A)=0.9,\quad P(B|\overline{A})=0.1,$$

- 于是, 用公式 (6.6) 得到

$$\begin{aligned}P(A|B)=&\frac{P(A)P(B|A)}{P(A)P(B|A)+P(\overline{A})P(B|\overline{A})}\\=&\frac{0.001\times 0.9}{0.001\times 0.9+0.999\times 0.1}\\=&\frac{9}{9+999}=0.0089<1\%.\end{aligned}$$

- 没有病的概率 $P(\overline{A}|B)=0.9911>99\%$.
- 造成这个结果的原因是发病率较低和诊断的准确性不够高.
- 如果甲复查时又被诊断有病, 则他的确有病的概率将增加到 7.5%.
- 如果人群的发病率不变, 诊断的准确率提高到 99%, 可以计算出 $P(A|B)=9.02\%$.

例 6.5(吸烟与肺癌问题)

- **例 6.5** (吸烟与肺癌问题) 1950 年某地区曾对 50-60 岁的男性公民进行调查. 肺癌病人中吸烟的比例是 99.7%, 无肺癌人中吸烟的比例是 95.8%. 如果整个人群的发病率是 $p=10^{-4}$, 求吸烟人群中的肺癌发病率和不吸烟人群中的肺癌发病率.
- **解** 引入 $A=$ 有肺癌, $B=$ 吸烟, 则

$$\begin{aligned}P(A)=&10^{-4},\\P(B|A)=&99.7\%,\\P(B|\overline{A})=&95.8\%.\end{aligned}$$

- 利用公式 (6.6) 得到:

$$\begin{aligned}P(A|B)=&\frac{P(A)P(B|A)}{P(A)P(B|A)+P(\overline{A})P(B|\overline{A})}\\=&\frac{10^{-4}\times 99.7\%}{10^{-4}\times 99.7\%+(1-10^{-4})\times 95.8\%}\\=&1.0407\times 10^{-4}.\\P(A|\overline{B})=&\frac{P(A)P(\overline{B}|A)}{P(A)P(\overline{B}|A)+P(\overline{A})P(\overline{B}|\overline{A})}\\=&\frac{10^{-4}\times(1-99.7\%)}{10^{-4}\times(1-99.7\%)+(1-10^{-4})\times(1-95.8\%)}\\=&7.1438\times 10^{-6}.\end{aligned}$$

- 于是,

$$\frac{\text{吸烟人群的发病率}}{\text{不吸烟人群的发病率}} = \frac{P(A|B)}{P(A|\overline{B})} = 14.57.$$

- 结论: 吸烟人群的肺癌发病率是不吸烟人群的肺癌发病率的 14.57 倍.

例 6.6(肇事车判定)

- **例 6.6** 某城市夏利牌出租车占 85%, 富康牌出租车占 15%. 这两种出租车都是红色, 富康出租车略大一些, 每辆车肇事的概率相同.

- 在一次出租车的交通肇事逃逸案件中, 有证人指证富康车肇事. 为了确定是否富康车肇事, 在肇事地点和相似的能见度下警方对证人辨别出租车的能力进行了测验, 发现证人正确识别富康车的概率是 90%, 正确识别夏利车的概率是 80%.

- 如果证人没有撒谎, 求富康车肇事的概率.

- **解:** 用 A 表示证人看见富康车肇事, 用 B 表示富康车肇事, 则 $\overline{B}$ 表示夏利车肇事, 并且

- $$P(B) = 0.15,\ P(A|B) = 0.9,\ P(A|\overline{B}) = 1 - 0.8.$$

- 要求的概率是 $P(B|A)$. 用 Bayes 公式得到

$$\begin{aligned} P(B|A) =& \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\overline{B})P(\overline{B})} \\ =& \frac{0.9 \times 0.15}{0.9 \times 0.15 + (1-0.8) \times 0.85} \\ =& 44.26\%. \end{aligned}$$

- 这个概率看起来很小, 但是在没有证人的情况下, 富康车肇事的概率更小, 是 15%.

1.7 概率与频率

概率与频率

- 古典概型只对等可能的情况定义了概率, 为了能够描述更复杂的试验, 很多学者使用概率的频率定义.