

习题课 1

罗雁天

March 21, 2019

1. 考虑医疗诊断问题，假设对于某种疾病，诊断的正确率为 p 。也就是说，如果就诊者确实患有该病，则医生能够以概率 p 做出准确诊断；如果就诊者实际没有患该病，医生做出正确判断的概率也是 p 。假设疾病自身的发病率是 q 。现已知某就诊者被医生诊断为患病，则其实际患该病的概率是多少？从中能得到什么结论呢？

解. 设事件 A 表示就诊者实际患病， D 表示就诊者被诊断为患病，那么由 Bayesian 公式：

$$P(A|D) = \frac{P(D|A)P(A)}{P(D)} = \frac{P(D|A)P(A)}{P(D|A)P(A) + P(D|A^c)P(A^c)} \quad (1)$$

带入 $P(A) = q, P(A^c) = 1 - q, P(D|A) = p, P(D|A^c) = 1 - p$ 得：

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

同理也可以得到：

$$P(A|D^c) = \frac{(1-p)q}{(1-p)q + p(1-q)} \quad (3)$$

下面我们通过一些有趣的计算来进一步认识这一问题。如果疾病的发病率很低 ($q = 0.01$)，医生的医术值得信任，诊断的正确率很高 ($p = 0.99$)，那么在诊断患病的条件下，我们有：

$$P(A|D) = \frac{0.99 \times 0.01}{0.99 \times 0.01 + 0.01 \times 0.99} = \frac{1}{2} \quad (4)$$

也就是说，真实患病的概率只有 50%。不难发现，条件概率的计算中，疾病本身较低的发病率使得医生判断的准确性大打折扣。尝试将发病率 q 升高至 0.5(例如普通感冒)，那么在医生医术保持不变的情况下，有：

$$P(A|D) = \frac{0.99 \times 0.5}{0.99 \times 0.5 + 0.01 \times 0.5} = 0.99 \quad (5)$$

可见，对于普通常见病，医生的水平能够得到充分的体现。但如果进一步降低发病率 ($q = 0.001$)，即所谓“疑难杂症”，如果维持医生水平不变，那么：

$$P(A|D) = \frac{0.99 \times 0.001}{0.99 \times 0.001 + 0.01 \times 0.999} = 0.09 \quad (6)$$

诊断的实际准确率连 10% 都不到，误报（实际没病，诊断有病）的概率达到了 90%。看起来，即使是非常称职的医生，当面临疑难杂症的时候也难免误报，患者应给予充分的理解，并在得知诊断结果之后保持冷静，争取用复诊的方法进一步确定是否患病。另一方面：

$$P(A|D^c) = \frac{0.01 \times 0.001}{0.01 \times 0.001 + 0.99 \times 0.999} \approx 10^{-5} \ll 1 \quad (7)$$

这说明，尽管误报的概率很高，但是漏报（实际患病，诊断无病）的概率却很低。所以，如果医生真的非常称职且水平很高，那么漏报性质的误诊率是能够充分降低的，即使面对的是十分罕见的疑难杂症。 \square

2. (匹配问题) n 个人随机地选取帽子，试问至少有一人戴上了自己帽子的概率是多少。

解. 设 B_k 表示第 k 个人戴对的所有可能结果构成的集合。则至少有一人戴对的所有可能结果可以描述为：

$$B_1 \cup B_2 \cup \cdots \cup B_n \quad (8)$$

我们的任务是计算这个集合的概率。由于：

$$P(B_k) = \frac{(n-1)!}{n!} = \frac{1}{n}, P(B_k \cap B_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}, \cdots \quad (9)$$

所以根据容斥原理，我们有：

$$\begin{aligned} P(B_1 \cup B_2 \cup \cdots \cup B_n) &= \sum_{k=1}^n P(B_k) - \sum_{j < k} P(B_j \cap B_k) + \sum_{i < j < k} P(B_i \cap B_j \cap B_k) \\ &\quad - \cdots + (-1)^{n+1} P(B_1 \cap B_2 \cap \cdots \cap B_n) \\ &= n \times \frac{1}{n} - \binom{n}{2} \frac{(n-2)!}{n!} + \binom{n}{3} \frac{(n-3)!}{n!} - \cdots + (-1)^{n+1} \frac{1}{n!} \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n+1} \frac{1}{n!} \\ &= \sum_{k=1}^n \frac{(-1)^{k+1}}{k!} \end{aligned} \quad (10)$$

即，至少有一人带上自己帽子的概率为 $\sum_{k=1}^n \frac{(-1)^{k+1}}{k!}$ \square

3. (匹配问题 2) 10 对夫妇坐一圈，计算所有的妻子都不坐在她丈夫旁边的概率

解. 令 $E_i (i = 1, 2, \dots, 10)$ 表示第 i 对夫妇坐在一起，因此，所求概率为：

$$P = 1 - P(E_1 \cup E_2 \cup \cdots \cup E_{10}) \quad (11)$$

根据容斥原理：

$$P(E_1 \cup E_2 \cup \cdots \cup E_{10}) = \sum_{i=1}^{10} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} E_{i_2}) + \cdots - P(E_1 E_2 \cdots E_{10}) \quad (12)$$

为了计算式子中每一项的概率，我们考虑通式 $P(E_{i_1}E_{i_2}\cdots E_{i_n})$ 。先注意到 20 个人坐一圈一共有 $19!$ 种可能的方式。对于指定的 n 对夫妇，在排位时使用“捆绑法”，这样在排位时，一共有 $20-n$ 个对象，在圆桌上一共有 $(20-n-1)! = (19-n)!$ 种排法。当排位确定之后，这 n 对夫妇之间又有排位问题，是男左女右还是男右女左，于是一共有 $2^n(19-n)!$ 种排法。因此我们得到：

$$P(E_{i_1}E_{i_2}\cdots E_{i_n}) = \frac{2^n(19-n)!}{19!} \quad (13)$$

带入式 (12) 中可以得到：

$$\begin{aligned} P(E_1 \cup E_2 \cup \cdots \cup E_{10}) &= \binom{10}{1} 2^1 \frac{18!}{19!} - \binom{10}{2} 2^2 \frac{17!}{19!} + \cdots - \binom{10}{10} 2^{10} \frac{9!}{19!} \\ &= \sum_{i=1}^{10} \binom{10}{i} 2^i \frac{(19-i)!}{19!} \end{aligned} \quad (14)$$

所以我们可以得到所有的妻子都不坐在她丈夫旁边的概率为：

$$P = 1 - \sum_{i=1}^{10} \binom{10}{i} 2^i \frac{(19-i)!}{19!} \approx 0.3395 \quad (15)$$

□

4. 重新考虑匹配问题。 n 个人随机地选取帽子，试问恰好有 k 人戴上了自己帽子的概率是多少。

解. 考虑戴上了自己帽子的 k 个人，第一个人戴对的概率是 $1/n$ ；第二个人受到了第一个人的条件限制，戴对的概率是 $1/(n-1)$ ，以此类推，得到

$$P(A_1) = \frac{1}{n}, P(A_2|A_1) = \frac{1}{n-1}, P(A_3|A_2A_1) = \frac{1}{n-2}, \cdots \quad (16)$$

因此：

$$P(A_1A_2\cdots A_k) = P(A_1)P(A_2|A_1)P(A_3|A_2A_1)\cdots P(A_k|A_{k-1}\cdots A_1) = \frac{(n-k)!}{n!} \quad (17)$$

另一方面，在 k 个人戴对的情况下，剩下 $n-k$ 个人都没有戴对，根据之前的结论，我们有：

$$P(B_1B_2\cdots B_{n-k}|A_1A_2\cdots A_k) = \sum_{j=0}^{n-k} \frac{(-1)^j}{j!} \quad (18)$$

则有：

$$P(A_1A_2\cdots A_kB_1B_2\cdots B_{n-k}) = \frac{(n-k)!}{n!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!} \quad (19)$$

考虑到戴对帽子的 k 个人的任意性，最终所求概率为：

$$P = \binom{n}{k} P(A_1A_2\cdots A_kB_1B_2\cdots B_{n-k}) = \frac{1}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!} \quad (20)$$

□

5. 某罪犯在犯罪现场留下了一些 DNA，法医研究后注意到能够辨认的只有 5 对，而且每个无罪的人，与这 5 对相匹配的概率为 10^{-5} ，律师认为罪犯就是该城镇 1 000 000 个居民之一。在过去 10 年内，该城镇有 10000 人刑满释放，他们的 DNA 资料都记录在案，律师认为这 10000 个有犯罪前科的人犯此罪的概率为 α ，而其余 990000 个居民中每个人犯此罪的概率为 β ，其中 $\alpha = c\beta$ 。(即他认为最近 10 年内释放的有犯罪前科的人作案的可能性是其他人的 c 倍。) 将 DNA 分析结果同这 10000 个有犯罪前科的人的数据文档对比后，发现只有 AJ 琼斯的 DNA 符合。假设律师关于 α 和 β 的关系是准确的，AJ 作案的可能性有多大？

解. 首先，概率之和必等于 1，我们有：

$$1 = 10000\alpha + 990000\beta = (10000c + 990000)\beta \quad (21)$$

因此：

$$\beta = \frac{1}{10000c + 990000}, \alpha = \frac{c}{10000c + 990000} \quad (22)$$

令 G 表示事件“AJ 为作案者”，令 M 表示事件“AJ 是这 10000 人中唯一与现场 DNA 相匹配的人”。那么：

$$P(G|M) = \frac{P(GM)}{P(M)} = \frac{P(M|G)P(G)}{P(M|G)P(G) + P(M|G^c)P(G^c)} \quad (23)$$

如果 AJ 为作案者，那么其他 9999 个人都不是作案者，那么我们有：

$$P(M|G) = (1 - 10^{-5})^{9999}, P(G) = \alpha \quad (24)$$

如果 AJ 不是作案者，若是他的 DNA 唯一匹配，那么其他 9999 个有前科的人一定不是作案者且都不是匹配者，因此作案者便出现在没有前科的人中。令 $C = \{\text{除 AJ 外，其他有前科的人都不是作案者}\}$ ，此时我们有：

$$P(C|G^c) = \frac{P(CG^c)}{G^c} = \frac{1 - 10000\alpha}{1 - \alpha} \quad (25)$$

同样，在除 AJ 外，其他有前科的人不是作案者的前提下，这 9999 个人与现场 DNA 不匹配的概率为 $(1 - 10^{-5})^{9999}$ ，所以

$$P(M|G^c) = 10^{-5}(1 - 10^{-5})^{9999} \frac{1 - 10000\alpha}{1 - \alpha}, P(G^c) = 1 - \alpha \quad (26)$$

带入得：

$$P(G|M) = \frac{\alpha}{\alpha + 10^{-5}(1 - 10000\alpha)} = \frac{1}{0.9 + \frac{10^{-5}}{\alpha}} \quad (27)$$

下面我们同样进行一些有趣的计算。如果律师认为任一有犯罪前科的人作案的可能性是没有犯罪前科的人的 100 倍 (即 $c=100$)，那么 $\alpha = 1/19900$ ，且：

$$P(G|M) = \frac{1}{1.099} \approx 0.9099 \quad (28)$$

如果律师最初认为 $c=10$, 那么 $\alpha = 1/109000$, 且:

$$P(G|M) = \frac{1}{1.99} \approx 0.5025 \quad (29)$$

如果律师最初认为 $c=1$, 即认为有犯罪前科的人作案的可能性与镇里其他人是相同的, 那么 $\alpha = 10^{-6}$, 且:

$$P(G|M) = \frac{1}{10.9} \approx 0.0917 \quad (30)$$

□

6. (选择题设置选项问题) 考试题目中有一道选择题, 学生会做的概率是 p 。如果会做, 那么做对的概率为 q , 如果不会做, 那么学生将会在 n 个答案中随机选取。假如判卷时发现某学生给出的答案是正确的, 那么该生的确会做这道题的概率是多少呢? 从中可以得到什么结论呢?

解. 设事件 A 表示学生的确会做该题目, 事件 C 表示学生给出了正确答案, 那么根据 Bayesian 公式:

$$P(A|C) = \frac{P(C|A)P(A)}{P(C|A)P(A) + P(C|A^c)P(A^c)} \quad (31)$$

带入 $P(C|A) = q, P(C|A^c) = 1/n$ 得:

$$P(A|C) = \frac{pq}{pq + (1-p)/n} \quad (32)$$

假设 $q = 1$ (学生们都非常细心, 会做的题目都不会错), 那么:

$$P(A|C) = \frac{np}{1 + (n-1)p} \quad (33)$$

在这种情况下如果令 $P(A|C) = \frac{1}{n}$ (即“瞎蒙”和“会做”的概率相等), 那么有:

$$p = \frac{1}{n^2 - n + 1} \quad (34)$$

这说明, 选择题的选项设置不能太少, 假如 $n = 2$, 那么只要题目稍难 ($1/3$ 的学生会做, $p=1/3$), 那么答案正确的情况下, 会做和随机选取的概率一样, 即“瞎蒙”和“会做”没有分别。另一方面, 题目不能出的太难, 以 $n = 4$ (通常的选择题) 为例, 如果会做的同学少于 $1/13 \approx 7\%$, 那么单纯从卷面答案上仍无法区分“瞎蒙”和“会做”。所以, 良好的选择题设计要经过认真的考虑。 □

7. 有 3 个付费公用电话, 每次打电话需投币 1 元, 但是投币后, A 电话始终正常工作, B 电话无法正常工作, C 电话以 $1/2$ 的概率正常工作。三部电话都不会退币。如果你只有 2 元, 那么采用什么样的策略, 能够让你成功打出电话的概率尽可能大呢? 此时打通电话的概率为多少?

解. 首先第一次随机选择一个电话进行拨打, 如果成功那么试验结束。第一次实验成功的概率为:

$$p_1 = \frac{1}{3} \times 1 + \frac{1}{3} \times \frac{1}{2} = \frac{1}{2} \quad (35)$$

若失败, 则该电话为 B 的概率为:

$$p_B = \frac{1 \times \frac{1}{3}}{\frac{1}{3} \times (0 + \frac{1}{2} + 1)} = \frac{2}{3} \quad (36)$$

那么, 该电话为 C 的概率为 $1/3$. 如果第二次继续选择该电话, 那么打通电话的概率为:

$$P_1 = \frac{2}{3} \times 0 + \frac{1}{3} \times \frac{1}{2} = \frac{1}{6} \quad (37)$$

如果随机选择另外一部电话, 则打通电话的概率为:

$$P_2 = \frac{2}{3} \left(\frac{1}{2} \times 1 + \frac{1}{2} \times \frac{1}{2} \right) + \frac{1}{3} \left(\frac{1}{2} \times 0 + \frac{1}{2} \times 1 \right) = \frac{2}{3} \quad (38)$$

由于 $P_1 < P_2$, 因此我们的最优策略为: 第一次随机选择一个电话拨打, 若成功则停止, 若失败, 则第二次选择另外两台中的一台进行拨打。在此策略下, 最终打通电话的概率为:

$$P = p_1 + (1 - p_1) \times P_2 = \frac{5}{6} \quad (39)$$

□