详细解答

STEP II 1987，第 16 题

概率与统计 • 贝叶斯定理

STEP II 1987，第 16 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/87-s2-q16

关键思路

这是贝叶斯定理（Bayes' theorem）最朴素的形式。两个“是”的报告可以有两种来源：说法属实而两位朋友都说真话，或者说法不实而两人都说谎。我们按每条途径的先验概率（prior probability）和出现这对报告的概率为它加权，然后问总数中有多大比例来自“属实”这条途径。第二部分的做法相同。唯一的新要点是，朋友的回答只涉及 A，所以我们必须按 A 是否在雪柜里，把三种可能情况分类。

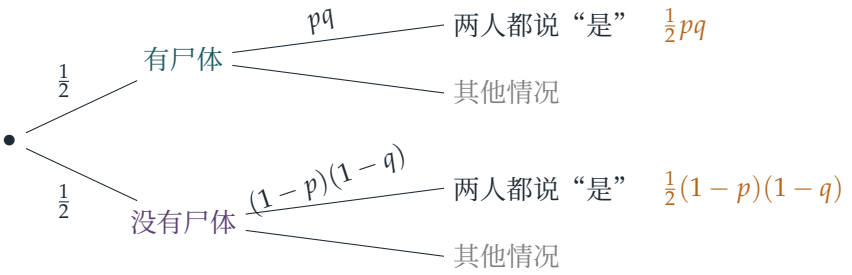
解答

第一个估计

设 F 为雪柜里有尸体这一事件，Y 为 P 和 Q 都说有尸体这一事件。今天早上 $\Pr(F) = \Pr(F') = \frac{1}{2}$。

我们把每位朋友看作回答一条是非题：说真话就是给出正确答案，说谎就是给出错误答案。朋友是否诚实与雪柜里有甚么无关，而且 P 和 Q 的行为互相独立（independent）。所以：

- 若有尸体，两人都说“是”当且仅当两人都说真话，所以 $\Pr(Y \mid F) = pq$；
- 若没有尸体，两人都说“是”当且仅当两人都说谎，所以 $\Pr(Y \mid F') = (1-p)(1-q)$。



根据条件概率（conditional probability）的定义（贝叶斯定理），

$$\begin{aligned}\Pr(F \mid Y) &= \frac{\Pr(F)\Pr(Y \mid F)}{\Pr(F)\Pr(Y \mid F) + \Pr(F')\Pr(Y \mid F')} = \frac{\frac{1}{2}pq}{\frac{1}{2}pq + \frac{1}{2}(1-p)(1-q)} \\ &= \frac{pq}{pq + 1 - p - q + pq} = \frac{pq}{1 - p - q + 2pq},\end{aligned}$$

这就是所求的估计。由于先验概率是平分的，因子 $\frac{1}{2}$ 互相抵消。

> ? **想一想** 为甚么答案不就是 pq，即两人都说真话的概率?

因为我们已经知道一些信息：他们的说法一致。恰好一人说谎的结果现在已不可能出现，因为那会产生一个“是”和一个“否”。我们必须在剩下的情况上重新按比例调整。在两条剩下的途径中，“两人都说真话”的权重是 pq，“两人都说谎”的权重是 $(1-p)(1-q)$。除以两者之和，正是这个重新调整。（先验概率 $\frac{1}{2}$ 也有作用。若先验概率不平分，它们就不会抵消。）

第二个估计

现在肯定有尸体。三个等可能的情况，各自的先验概率为 $\frac{1}{3}$，分别是

$$\text{只有 } A, \qquad \text{只有 } B, \qquad A \text{ 和 } B \text{ 都在.}$$

设 Y_A 为两位朋友都说 A 在雪柜里这一事件。按照与之前相同的模型，他们的回答只取决于 A 是否在雪柜里：

- 在“只有 A”和“两者都在”的情况下，A 在雪柜里，所以每种情况都有 $\Pr(Y_A) = pq$；
- 在“只有 B”的情况下，A 不在，所以 $\Pr(Y_A) = (1-p)(1-q)$。

因此

$$\begin{aligned}\Pr(\text{两者都在} \mid Y_A) &= \frac{\frac{1}{3}pq}{\frac{1}{3}pq + \frac{1}{3}(1-p)(1-q) + \frac{1}{3}pq} \\ &= \frac{pq}{2pq + 1 - p - q + pq} = \frac{pq}{1 - p - q + 3pq}.\end{aligned}$$

留意这些报告并没有直接告诉我们任何关于 B 的信息。它们唯一的作用，是改变 A 在场的概率。根据第一部分（以先验概率 $\frac{2}{3}$ 代替 $\frac{1}{2}$），这个概率变成 $2pq/(1-p-q+3pq)$。已知 A 在场时，“只有 A”和“两者都在”仍然等可能，所以把它减半。这样得出相同的答案。

验算

- 若 $p=q=\frac{1}{2}$，朋友的话毫无用处，答案回到先验概率：$\frac{1}{4}/\frac{1}{2}=\frac{1}{2}$ 及 $\frac{1}{4}/\frac{3}{4}=\frac{1}{3}$。
- 若 $p=q=1$，朋友绝不出错，答案变成 1（有尸体）和 $\frac{1}{2}$（A 在，所以 B 是否也在是一半一半）。
- 当 $p=q=\frac{3}{4}$ 时，答案是 $\frac{9}{10}$ 和 $\frac{9}{19}$。对 200 组随机的 (p,q)，逐一列举所有状态与说真话情况组合的精确计算均与答案吻合。以 $p=0.7$、$q=0.6$ 进行 4×10^5 次试验的模拟，第二部分得出 0.4375，与 $\frac{0.42}{0.96}=0.4375$ 相符。

常见错误

- 答 pq，忘记了“两人都说谎”同样会产生两个“是”。
- 在第二部分，分子只放“两者都在”的情况，却在分母中漏掉“只有 A”的情况。这两种情况都与 A 在场相符。

STEP 之外

这就是贝叶斯更新（Bayesian updating）。先验信念乘以证据的似然（likelihood），然后归一化。一个简洁的写法是胜算（odds）形式：后验胜算等于先验胜算乘以似然比（likelihood ratio）$\frac{pq}{(1-p)(1-q)}$。这样写就清楚显示，互相独立的证人只需把各自的比率相乘。同样的计算是医学检测、垃圾邮件过滤器以及法庭上综合证据的基础。Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》）有一章精彩地讲解条件概率，收录了许多正是这类风格的趣题。

Beyond STEP 详细解答：https://beyondstep.keithmcnulty.org/zh-hans/。解答文字采用 CC0 1.0；题目本身版权属 Cambridge University Press & Assessment 所有（©），本文件不予转载。由机器从英文翻译而成；尚未经审阅。