详细解答

STEP II 1999，第 12 题

概率与统计 · 贝叶斯定理

> **STEP II 1999，第 12 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/99-s2-q12

关键思路

这是一道保留了参数的贝叶斯定理（Bayes' theorem）题目。只要用全概率公式（law of total probability）求出 $\mathrm{P}(S)$，两个条件概率（conditional probability）都是 p 的简单有理函数（rational function），而且各自在容许的区间上都是单调（monotonic）的。因此，每位检查员的做法中所谓的“最大值”，都落在区间 $0 \le p \le \frac{1}{3}$ 的端点上；而真正有趣的地方，在于这对单凭一次观察作估计有甚么启示。

解答

条件概率

设 A、B、C 分别为随机抽出的一块晶片由 A、B、C 制造的事件，并设 S 为该晶片完好的事件。这三个事件构成样本空间（sample space）的一个分割（partition），其中

$$\mathrm{P}(A) = 2p, \quad \mathrm{P}(B) = p, \quad \mathrm{P}(C) = 1 - 3p,$$
$$\mathrm{P}(S \mid A) = \tfrac{7}{10}, \quad \mathrm{P}(S \mid B) = \tfrac{8}{10}, \quad \mathrm{P}(S \mid C) = \tfrac{9}{10}.$$

根据全概率公式，

$$\begin{aligned}\mathrm{P}(S) &= 2p \cdot \tfrac{7}{10} + p \cdot \tfrac{8}{10} + (1 - 3p) \cdot \tfrac{9}{10} \\ &= \tfrac{1}{10}(14p + 8p + 9 - 27p) = \tfrac{1}{10}(9 - 5p).\end{aligned}$$

留意在整个 $0 \le p \le \frac{1}{3}$ 的范围内都有 $9 - 5p > 0$，所以可以用它作除数。根据贝叶斯定理，

$$\mathrm{P}(A \mid S) = \frac{\mathrm{P}(S \mid A)\,\mathrm{P}(A)}{\mathrm{P}(S)} = \frac{\frac{14p}{10}}{\frac{9-5p}{10}} = \frac{14p}{9 - 5p},$$

同理

$$\mathrm{P}(C \mid S) = \frac{\frac{9}{10}(1-3p)}{\frac{1}{10}(9-5p)} = \frac{9(1-3p)}{9-5p} = \frac{9-27p}{9-5p}.$$

第一位检查员的估计

当 p 由 0 增加至 $\frac{1}{3}$ 时，分子 $14p$ 递增，而正的分母 $9-5p$ 递减，所以 $\mathrm{P}(A \mid S)$ 严格递增（strictly increasing）。（等价地说，它的导数是 $126/(9-5p)^2 > 0$。）因此它的最大值出现在右端点 $p = \frac{1}{3}$，其值为

$$\frac{14/3}{9-5/3} = \frac{14/3}{22/3} = \frac{7}{11}.$$

所以第一位检查员的估计是 $\hat{p} = \frac{1}{3}$。

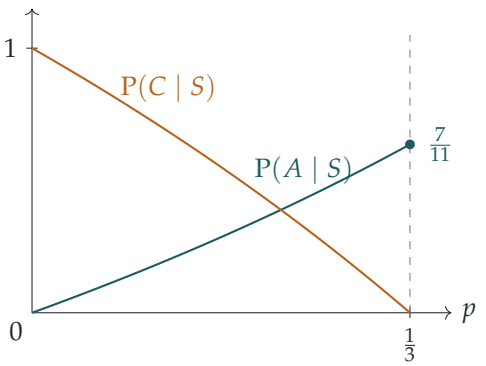
第二位检查员的估计

> ? **想一想** $\mathrm{P}(C \mid S)$ 是递增还是递减？随着 p 增大，它的分子和分母都在减小。

当分子和分母朝同一方向变化时，我们不能一眼看出答案，所以要用商法则（quotient rule）求导：

$$\frac{\mathrm{d}}{\mathrm{d}p}\,\frac{9-27p}{9-5p} = \frac{-27(9-5p)+5(9-27p)}{(9-5p)^2} = \frac{-243+135p+45-135p}{(9-5p)^2} = \frac{-198}{(9-5p)^2} < 0.$$

所以 $\mathrm{P}(C \mid S)$ 在 $[0, \frac{1}{3}]$ 上严格递减，它的最大值出现在 $p = 0$，其值为 1。第二位检查员的估计是 $\hat{p} = 0$。



评论

两位检查员用的是同一个方法，各自依据的单次观察也同样有效，却得出容许范围内两个相反的极端：$\hat{p} = \frac{1}{3}$（完全没有晶片来自 C）和 $\hat{p} = 0$（所有晶片都来自 C）。每个估计其实只是“令我这块晶片的制造商尽可能常见的 p 值”，而只有一个样本时，必然会得出

这样的结果。第二个估计甚至与第一位检查员手上的晶片互相矛盾，因为 $p = 0$ 表示 A 根本不制造晶片。两个估计都不可信：一次观察所含的信息实在太少，不足以估计 p，我们需要更大的样本（连同来自每个制造商的晶片数目）。

还有一点值得留意：晶片是否完好，几乎不起任何作用。“完好且来自 A”的似然（likelihood）是 $\frac{7}{10} \cdot 2p$，它取得最大值的位置与 $2p$ 本身相同；因子 $\frac{7}{10}$ 与 p 无关。

验算

- 当 $p = 0$ 时，所有晶片都来自 C，所以 $\mathrm{P}(C \mid S)$ 应为 1，而 $\mathrm{P}(A \mid S)$ 应为 0；两条公式都与此吻合。
- 同理可得 $\mathrm{P}(B \mid S) = 8p/(9 - 5p)$，而三个条件概率之和为

$$\frac{14p + 8p + 9 - 27p}{9 - 5p} = 1.$$

- 当 $p = 0.2$ 时，公式给出 $\mathrm{P}(A \mid S) = 0.35$ 和 $\mathrm{P}(C \mid S) = 0.45$；以 400 000 块晶片进行模拟，得出 0.3505 和 0.4505。

常见错误

- 令导数等于零，然后寻找转向点（turning point）。这里根本没有转向点：两个最大值都在 $[0, \frac{1}{3}]$ 的端点上。
- 忘记 $\mathrm{P}(C \mid S)$ 需要仔细论证：分子和分母都在减小，所以单说“分子变小了”并不足够。

STEP 之外

两位检查员所做的，实际上就是最大似然估计（maximum likelihood estimation）：选出令他们所见的结果概率最大的参数值。只有一次观察时，最大似然估计值往往落在参数空间（parameter space）的边界上，这提醒我们：它需要大量数据才会表现良好。贝叶斯学派（Bayesian）的做法则是先为 p 设置一个先验分布（prior distribution），然后每观察一块晶片就更新一次，这样便不会单凭一次观察而得出如此极端的结论。Blitzstein 与 Hwang 所著的 *Introduction to Probability*（概率论入门教材）对这两种观点都有讨论。

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。