详细解答

STEP I 2000，第 13 题

概率与统计 • 条件概率

STEP I 2000，第 13 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/00-s1-q13

关键思路

整个问题归根究柢就是追踪父母各自传下来的一个基因。在 (i) 中，两个 AB 型父母所生的每个孩子，是 AA 或 BB 的概率（probability）同样那么小，而四个孩子互相独立（independent），所以我们只需数出排列方式的数目，再把概率相乘。第 (ii) 部分是“逆向”的条件概率（conditional probability）：我们知道结果（孩子），要找的是原因（父亲）。这正是贝叶斯定理（Bayes' theorem），而按父亲三种可能的基因型（genotype）画出树形图（tree diagram），就能把计算整理得井井有条。

解答

第 (i) 部分

父母各自都是 AB，所以各自以概率 $\frac{1}{2}$ 传下 A 或 B，而且父母二人互相独立。对单独一个孩子而言，

$$\mathrm{P}(AA) = \tfrac{1}{2}\cdot\tfrac{1}{2} = \tfrac{1}{4}, \qquad \mathrm{P}(BB) = \tfrac{1}{4}, \qquad \mathrm{P}(AB) = \tfrac{1}{2},$$

其中最后一个已把两种次序都计算在内（A 来自母亲而 B 来自父亲，或者反过来）。

不同孩子的遗传互相独立。四个孩子中两个是 AA、两个是 BB 的某一个特定次序，例如 AA, AA, BB, BB，其概率为 $\left(\frac{1}{4}\right)^4$。从四个孩子中选出哪两个是 AA 的方法数目是 $\binom{4}{2} = 6$。所以

$$\mathrm{P}(\text{两个}AA\text{ 和两个}BB) = 6\times\left(\tfrac{1}{4}\right)^4 = \frac{6}{256} = \frac{3}{128},$$

正如题目所求。

第 (ii) 部分

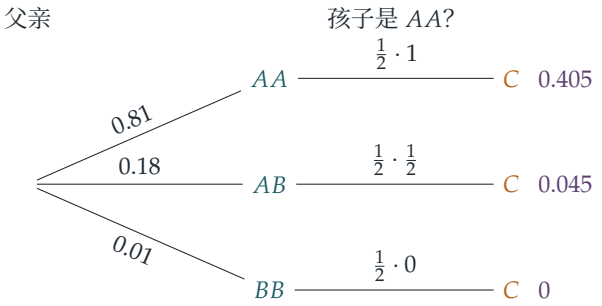
设 F_{AA}、F_{AB}、F_{BB} 分别为我父亲是 AA、AB、BB 的事件，并设 C 为我是 AA 的事件。我们把父亲视为从人口中随机抽出的一员，而且这个选取与母亲的基因型无关，所以

$$\mathrm{P}(F_{AA}) = 0.81, \qquad \mathrm{P}(F_{AB}) = 0.18, \qquad \mathrm{P}(F_{BB}) = 0.01.$$

> **想一想** 为甚么 0.18 还不是答案?

因为我们有额外的资料：我是 AA。这完全排除了父亲是 BB 的可能，同时令父亲是 AA 的可能性增加，因为 AA 型的父亲必定传下 A。所以我们必须以 C 为条件。

我要成为 AA，就需要从母亲得到一个 A，也从父亲得到一个 A。我母亲是 AB，所以不论父亲是哪一种基因型，她传下 A 的概率都是 $\frac{1}{2}$。父亲传下 A 的概率则视乎他的基因型，分别为 1、$\frac{1}{2}$ 或 0。树形图如下：



根据全概率公式（law of total probability），

$$\begin{aligned}\mathrm{P}(C) &= 0.81 \cdot \tfrac{1}{2} \cdot 1 + 0.18 \cdot \tfrac{1}{2} \cdot \tfrac{1}{2} + 0.01 \cdot \tfrac{1}{2} \cdot 0 \\ &= 0.405 + 0.045 + 0 = 0.45.\end{aligned}$$

再根据贝叶斯定理，

$$\mathrm{P}(F_{AB} \mid C) = \frac{\mathrm{P}(F_{AB})\,\mathrm{P}(C \mid F_{AB})}{\mathrm{P}(C)} = \frac{0.045}{0.45} = \frac{1}{10}.$$

所以我父亲是 AB 的概率是 $\frac{1}{10}$。

验算

- 在 (i) 中，四个孩子各种基因型数目组合的概率，总和应为 1；$\binom{4}{2}$ 这个因子，正是 $\left(\frac{1}{4} + \frac{1}{2} + \frac{1}{4}\right)^4$ 的多项式展开（multinomial expansion）中出现的那一个，而 $3/128 \approx 0.0234$ 就是其中一项。把八个基因的全部 2^8 种遗传方式逐一列举，所得的结果正是 $3/128$。

- 在 (ii) 中，答案由先验概率（prior）0.18 降至 0.1，这是合理的：父亲是 BB 的可能已被排除，而父亲是 AA 的可能性则有所增加。模拟 10^6 个家庭，所得结果为 0.0998。

常见错误

- 在 (i) 中漏掉因子 $\binom{4}{2} = 6$，得出 $1/256$。
- 在 (ii) 中答 0.18，或者计算了 $\mathrm{P}(C \mid F_{AB}) = \frac{1}{4}$，而不是 $\mathrm{P}(F_{AB} \mid C)$。
- 只在部分分支中漏掉母亲的因子 $\frac{1}{2}$：这个因子最终会约去，但前提是每一个分支都一致地使用它。

STEP 之外

人口中的比例 81%、18%、1% 正是 p^2、$2pq$、q^2，其中 $p = 0.9$、$q = 0.1$：这就是哈迪－温伯格平衡（Hardy–Weinberg equilibrium），在随机交配之下，它会一代一代地保持不变。从观察到的结果推断隐藏的原因，是贝叶斯统计学（Bayesian statistics）的核心，而同样的计算也是遗传咨询和亲子鉴定的基础。Blitzstein and Hwang, *Introduction to Probability*（概率论入门）对条件概率和贝叶斯定理有详尽的论述，并附有许多与本题类似的例子。

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。