詳細解答

STEP I 2000，第 13 題

概率與統計 • 條件概率

> **STEP I 2000，第 13 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/00-s1-q13

關鍵思路

整個問題歸根究柢就是追蹤父母各自傳下來的一個基因。在 (i) 中，兩個 AB 型父母所生的每個孩子，是 AA 或 BB 的概率（probability）同樣那麼小，而四個孩子互相獨立（independent），所以我們只需數出排列方式的數目，再把概率相乘。第 (ii) 部分是「逆向」的條件概率（conditional probability）：我們知道結果（孩子），要找的是原因（父親）。這正是貝葉斯定理（Bayes' theorem），而按父親三種可能的基因型（genotype）畫出樹形圖（tree diagram），就能把計算整理得井井有條。

解答

第 (i) 部分

父母各自都是 AB，所以各自以概率 $\frac{1}{2}$ 傳下 A 或 B，而且父母二人互相獨立。對單獨一個孩子而言，

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

其中最後一個已把兩種次序都計算在內（A 來自母親而 B 來自父親，或者反過來）。

不同孩子的遺傳互相獨立。四個孩子中兩個是 AA、兩個是 BB 的某一個特定次序，例如 AA, AA, BB, BB，其概率為 $\left(\frac{1}{4}\right)^4$。從四個孩子中選出哪兩個是 AA 的方法數目是 $\binom{4}{2} = 6$。所以

$$\mathrm{P}(\text{兩個}AA\text{ 和兩個}BB) = 6\times\left(\tfrac{1}{4}\right)^4 = \frac{6}{256} = \frac{3}{128},$$

正如題目所求。

第 (ii) 部分

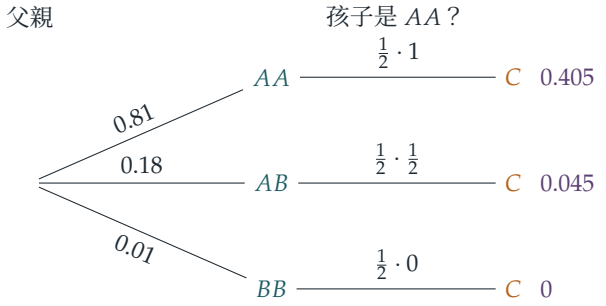
設 F_{AA}、F_{AB}、F_{BB} 分別為我父親是 AA、AB、BB 的事件，並設 C 為我是 AA 的事件。我們把父親視為從人口中隨機抽出的一員，而且這個選取與母親的基因型無關，所以

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

> ? **想一想** 為甚麼 0.18 還不是答案？

因為我們有額外的資料：我是 AA。這完全排除了父親是 BB 的可能，同時令父親是 AA 的可能性增加，因為 AA 型的父親必定傳下 A。所以我們必須以 C 為條件。

我要成為 AA，就需要從母親得到一個 A，也從父親得到一個 A。我母親是 AB，所以不論父親是哪一種基因型，她傳下 A 的概率都是 $\frac{1}{2}$。父親傳下 A 的概率則視乎他的基因型，分別為 1、$\frac{1}{2}$ 或 0。樹形圖如下：



根據全概率公式（law of total probability），

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

再根據貝葉斯定理，

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

所以我父親是 AB 的概率是 $\frac{1}{10}$。

驗算

- 在 (i) 中，四個孩子各種基因型數目組合的概率，總和應為 1；$\binom{4}{2}$ 這個因子，正是 $\left(\frac{1}{4} + \frac{1}{2} + \frac{1}{4}\right)^4$ 的多項式展開（multinomial expansion）中出現的那一個，而 $3/128 \approx 0.0234$ 就是其中一項。把八個基因的全部 2^8 種遺傳方式逐一列舉，所得的結果正是 $3/128$。

- 在 (ii) 中，答案由先驗概率（prior）0.18 降至 0.1，這是合理的：父親是 BB 的可能已被排除，而父親是 AA 的可能性則有所增加。模擬 10^6 個家庭，所得結果為 0.0998。

常見錯誤

- 在 (i) 中漏掉因子 $\binom{4}{2} = 6$，得出 $1/256$。
- 在 (ii) 中答 0.18，或者計算了 $\mathrm{P}(C \mid F_{AB}) = \frac{1}{4}$，而不是 $\mathrm{P}(F_{AB} \mid C)$。
- 只在部分分支中漏掉母親的因子 $\frac{1}{2}$：這個因子最終會約去，但前提是每一個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 $p = 0.9$、$q = 0.1$：這就是哈迪－溫伯格平衡（Hardy–Weinberg equilibrium），在隨機交配之下，它會一代一代地保持不變。從觀察到的結果推斷隱藏的原因，是貝葉斯統計學（Bayesian statistics）的核心，而同樣的計算也是遺傳諮詢和親子鑑定的基礎。Blitzstein and Hwang, *Introduction to Probability*（概率論入門）對條件概率和貝葉斯定理有詳盡的論述，並附有許多與本題類似的例子。

Beyond STEP 詳細解答：https://beyondstep.keithmcnulty.org/zh-hant/。解答文字採用 CC0 1.0；題目本身版權屬 Cambridge University Press & Assessment 所有（©），本文件不予轉載。由機器從英文翻譯而成；尚未經審閱。