解答例

STEP I 2000、第 13 問

確率・統計 ・ 条件付き確率

> **STEP I 2000、問 13**
>
> 先に進む前に、まず問題を読んでおこう。問題は STEP Support Programme の問題データベースにある：
>
> step.maths.org/questions/00-s1-q13

鍵となる考え方

すべては、両親それぞれから受け継ぐ遺伝子を一つずつ追跡することに帰着する。(i) では、AB の両親から生まれる子はそれぞれ同じ小さな確率で AA または BB となり、4 人の子は互いに独立（independent）であるから、並び方を数えて掛ければよい。(ii) は「逆向き」の条件付き確率（conditional probability）である。結果（子）がわかっていて、その原因（父親）を知りたいのである。これはベイズの定理（Bayes' theorem）であり、父親のとりうる 3 つの遺伝子型（genotype）についての樹形図（tree diagram）を描くと整理できる。

解答

小問 (i)

両親はともに AB であるから、それぞれ確率 $\frac{1}{2}$ で A または B を子に伝え、両親は互いに独立にふるまう。1 人の子について、

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

である。ただし最後のものは両方の順序（母親から A、父親から B を受け継ぐ場合と、その逆の場合）を数えている。

異なる子どうしは互いに独立に遺伝子を受け継ぐ。AA が 2 人、BB が 2 人となる 4 人の子の特定の並び、たとえば AA, AA, BB, BB の確率は $\left(\frac{1}{4}\right)^4$ である。4 人の子のうちどの 2 人が AA であるかを選ぶ方法は $\binom{4}{2} = 6$ 通りある。よって

$$\mathrm{P}(\text{2 人が}AA\text{、2 人が}BB) = 6 \times \left(\tfrac{1}{4}\right)^4 = \frac{6}{256} = \frac{3}{128},$$

となり、題意が示された。

小問 (ii)

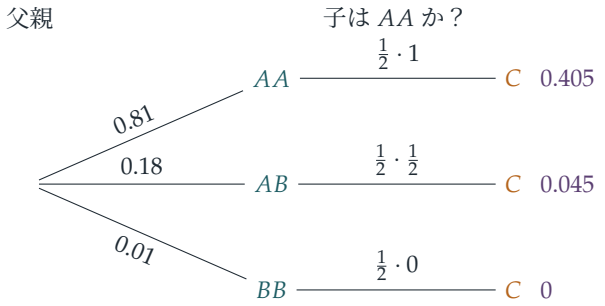
父親が AA、AB、BB であるという事象をそれぞれ F_{AA}、F_{AB}、F_{BB} とし、私が AA であるという事象を C とする。父親は母親の遺伝子型とは独立に選ばれた集団の無作為な一員であるとみなす。したがって

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

考えてみよう なぜ 0.18 がそのまま答えにならないのだろうか。

追加の情報、すなわち私が AA であるという情報があるからである。これにより父親が BB である可能性は完全に排除され、また AA の父親は必ず A を伝えるので、父親が AA である可能性は高くなる。したがって C を条件としなければならない。

私が AA であるためには、母親から A、父親から A を受け継ぐ必要がある。母親は AB であるから、父親が誰であっても確率 $\frac{1}{2}$ で A を伝える。父親は遺伝子型に応じて確率 1、$\frac{1}{2}$、0 で A を伝える。樹形図は次のとおりである。



全確率の法則（law of total probability）により、

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

次に、ベイズの定理により、

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

よって、父親が AB である確率は $\frac{1}{10}$ である。

検算

- (i) において、4 人の子の遺伝子型の個数についての確率の総和は 1 になるはずである。因子 $\binom{4}{2}$ は $\left(\frac{1}{4}+\frac{1}{2}+\frac{1}{4}\right)^4$ の多項展開（multinomial expansion）に現れるものと同じであり、$3/128 \approx 0.0234$ はその項の一つである。8 個の遺伝子の受け継ぎ方 2^8 通りをすべて正確に数え上げると $3/128$ が得られる。
- (ii) において、答えは事前確率（prior）0.18 から 0.1 に下がっている。BB の父親が排除され、AA の父親が有利になるのだから、そうなるべきである。10^6 家族のシミュレーションでは 0.0998 が得られた。

よくある間違い

- (i) で因子 $\binom{4}{2}=6$ を忘れて $1/256$ としてしまう。
- (ii) で 0.18 と答えてしまう、あるいは $\mathrm{P}(F_{AB} \mid C)$ ではなく $\mathrm{P}(C \mid F_{AB})=\frac{1}{4}$ を計算してしまう。
- 母親の因子 $\frac{1}{2}$ を一部の枝でだけ書き落とす。この因子は最終的に約分されて消えるが、それは一貫して用いた場合に限られる。

STEP のその先へ

集団における割合 81%、18%、1% は、$p=0.9$、$q=0.1$ としたときの p^2、$2pq$、q^2 である。これはハーディ・ワインベルク平衡（Hardy–Weinberg equilibrium）であり、無作為交配のもとでは世代から世代へと保たれる。観測された結果から隠れた原因を推論することはベイズ統計学（Bayesian statistics）の核心であり、同じ計算が遺伝カウンセリングや親子鑑定の基礎となっている。Blitzstein と Hwang による *Introduction to Probability*（確率論入門）は、条件付けとベイズの定理を、本問のような多くの例とともに丁寧に扱っている。

Beyond STEP 解答集：https://beyondstep.keithmcnulty.org/ja/。解答本文は CC0 1.0 である。問題そのものは © Cambridge University Press & Assessment に帰属し、ここには掲載していない。英語版から機械翻訳したものであり、まだ校閲を経ていない。