解答例

STEP I 2004、第 12 問

確率・統計 ・ ベイズの定理

> **STEP I 2004、問 12**
>
> 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/04-s1-q12

鍵となる考え方

グループ 1 から取り出した標本からは、チップがグループ 1 に入っているという条件のもとでそれが機械 A で作られた条件付き確率（conditional probability）がおおよそわかる。そこで、ベイズの定理（Bayes' theorem）を用いてこの条件付き確率を λ、p、q で表し、観測された割合と等しいとおいて λ について解く。選別に誤りがある場合に変わるのは、各機械のチップがグループ 1 に入る確率だけであり、解法はまったく同じである。

解答

工場の生産品から無作為にチップを 1 個選ぶ。それが機械 A で作られたという事象を A とし（したがって $\mathrm{P}(A) = \lambda$、$\mathrm{P}(A') = 1 - \lambda$ である）、それが完全品であるという事象を P、それがグループ 1 に入れられるという事象を G とする。問題文より $\mathrm{P}(P \mid A) = p$、$\mathrm{P}(P \mid A') = q$ である。

グループ 1 から無作為に取り出したチップとは、生産品から無作為に取り出したチップに G という条件をつけたものである。標本は大きいので、観測された割合 $\frac{2}{5}$ は $\mathrm{P}(A \mid G)$ のよい推定値であり、λ の推定値は $\mathrm{P}(A \mid G)$ をちょうど $\frac{2}{5}$ にする値である。

選別が完全な場合

このとき G は P と同じ事象である。ベイズの定理により、

$$\mathrm{P}(A \mid P) = \frac{\mathrm{P}(P \mid A)\,\mathrm{P}(A)}{\mathrm{P}(P \mid A)\,\mathrm{P}(A) + \mathrm{P}(P \mid A')\,\mathrm{P}(A')} = \frac{\lambda p}{\lambda p + (1-\lambda)q}.$$

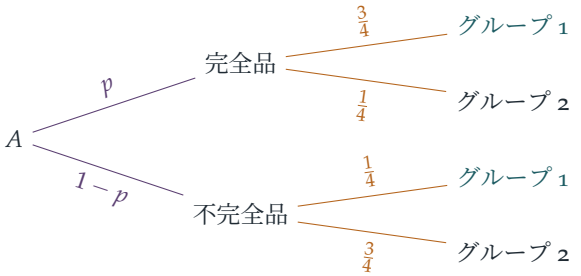
これを $\frac{2}{5}$ と等しいとおき、分母を払う（分母は正である）と、

$$5\lambda p = 2\lambda p + 2(1-\lambda)q$$
$$3\lambda p + 2\lambda q = 2q$$
$$\lambda = \frac{2q}{3p+2q},$$

となる。$3p+2q>0$ であるから、これが求める推定値である。

選別に誤りがある場合

このとき、チップがグループ 1 に入るのは、完全品で正しく選別される場合か、不完全品で誤って選別される場合のどちらかである。選別の誤りは、チップをどちらの機械が作ったかとは独立に起こると仮定する。つまり、チップが完全品かどうかが与えられれば、その選別のされ方は機械によらない。下の樹形図は機械 A のチップについての枝を示したものである。機械 B については p を q に置き換えればよい。



グループ 1 に至る 2 本の枝を足し合わせると、

$$\mathrm{P}(G \mid A) = \tfrac{3}{4}p + \tfrac{1}{4}(1-p) = \frac{1+2p}{4}, \qquad \mathrm{P}(G \mid A') = \tfrac{3}{4}q + \tfrac{1}{4}(1-q) = \frac{1+2q}{4}.$$

よって、再びベイズの定理により（因子 $\frac{1}{4}$ は約分される）、

$$\mathrm{P}(A \mid G) = \frac{\lambda(1+2p)}{\lambda(1+2p)+(1-\lambda)(1+2q)}.$$

考えてみよう 証拠は変わったのか？　$\frac{2}{5}$ をそのまま使ってよいのか？

よい。標本は実際にグループ 1 から取り出されたものであり、その $\frac{2}{5}$ は実際に機械 A で作られたものである。変わったのは、チップがどのようにしてグループ 1 に入った

かについてのモデルだけである。そこで、$\mathrm{P}(A \mid G)$ の新しい式を $\frac{2}{5}$ と等しいとおく。

$$\begin{aligned} 5\lambda(1+2p) &= 2\lambda(1+2p) + 2(1-\lambda)(1+2q) \\ \lambda\big[3(1+2p) + 2(1+2q)\big] &= 2(1+2q) \\ \lambda &= \frac{2(1+2q)}{5+6p+4q}. \end{aligned}$$

これが λ の新しい推定値である。これは、最初の答えにおいて p と q を、グループ 1 に入る新しい確率 $\frac{1+2p}{4}$ と $\frac{1+2q}{4}$ に置き換えたものになっていることに注意しよう。

検算

- $p = q$ ならば、品質によって 2 台の機械を区別することはできないので、グループ 1 は生産の割合をそのまま反映し、λ は $\frac{2}{5}$ になるはずである。実際、どちらの式もこの値を与える：$\frac{2p}{5p} = \frac{2}{5}$、$\frac{2(1+2p)}{5+10p} = \frac{2}{5}$。
- どちらの推定値も、分子が正で分母より小さいので、0 と 1 の間（両端を含まない）にある。
- $\lambda = 0.3$、$p = 0.6$、$q = 0.8$ として、選別に誤りがある場合の 4×10^5 個のチップのシミュレーションを行ったところ、グループ 1 のうち機械 A のものの割合は 0.2667 となり、公式による値 0.2661 とよく合った。シミュレーションで得た割合から公式を逆に解くと、$\lambda \approx 0.301$ が復元された。

よくある間違い

- $\frac{2}{5}$ を、条件付き確率 $\mathrm{P}(A \mid G)$ ではなく、$\mathrm{P}(G \mid A)$ や λp と等しいとおいてしまう。
- 後半で、誤ってグループ 1 に入れられた不完全品を忘れてしまう。つまり、$\frac{3}{4}p$ だけを用いてしまう。

STEP のその先へ

観測データに「ちょうど合う」ようにパラメータの値を選ぶ方法は**モーメント法**（method of moments）とよばれる。この問題では、それは大学の最初の統計学の講義で学ぶ最尤推定（maximum likelihood estimation）と一致する。誤りのある選別機は、測定誤差（誤分類、misclassification）の簡単なモデルである。これは医学検査において実際に重要な問題であり、偽陽性（false positive）と偽陰性（false negative）は、まさにこのようにして観測される割合を歪める。ベイズの定理と条件付けについてさら

に学ぶには、Blitzstein and Hwang, *Introduction to Probability*（確率論入門）を参照するとよい。

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。