解答例

# STEP II 2024、第 11 問

確率・統計 ・ グループ検査の期待値

> **STEP II 2024、問 11**
>
> 先に進む前に、まず問題を読んでおこう。問題は、OCR が配布している 2024 年の STEP 過去問一式（zip ファイル）のうち、STEP II の問題用紙にある：
>
> www.ocr.org.uk/Images/725753-2024-past-papers.zip

## 鍵となる考え方

検体をまとめて検査する方法（pooling）が得になるのは、グループ検査がたいてい陰性で返ってくる場合に限られる。1 人あたりの検査回数の期待値（expected value）は $1+\frac{1}{k}-(1-p)^k$ となる。これが 1 以下となるのはいつかを問うと、$(1-p)$ と $k^{-1/k}$ の比較に帰着する。小問 (i) で $x^{1/x}$ について問われているのはまさにこのためである。すなわち、$p$ が大きいときに最適なグループの大きさは、$k^{1/k}$ を最大にする整数なのである。

## 解答

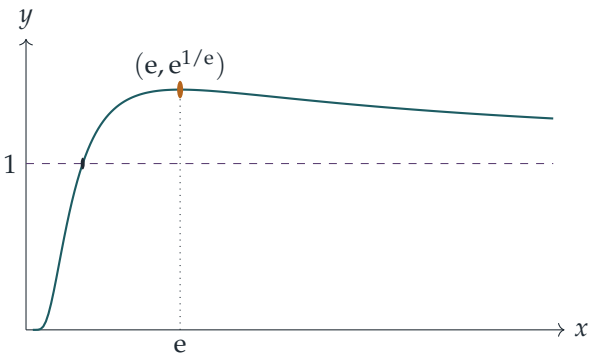
### 小問 (i)

$x>0$ に対して $y=x^{1/x}=\mathrm{e}^{(\ln x)/x}$ とおくと、$\ln y=\dfrac{\ln x}{x}$ である。微分すると

$$\frac{1}{y}\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{1-\ln x}{x^2}.$$

$y>0$ であるから、$\frac{\mathrm{d}y}{\mathrm{d}x}$ の符号は $1-\ln x$ の符号と一致する。すなわち $0<x<\mathrm{e}$ で正、$x=\mathrm{e}$ で 0、$x>\mathrm{e}$ で負である。したがって停留点（turning point）はただ 1 つであり、それは $(\mathrm{e},\mathrm{e}^{1/\mathrm{e}})\approx(2.72,\,1.44)$ における極大（maximum）である。

グラフの両端について調べる。$x\to 0^+$ のとき $(\ln x)/x\to-\infty$ であるから $y\to 0$ である。$x\to\infty$ のとき $(\ln x)/x\to 0$ であるから、$y$ は上から 1 に近づく（$x>1$ では $\ln y>0$ であるため）。曲線は点 $(1,1)$ を通る。




正の整数 $n$ について考える。この関数（function）は $(0, \mathrm{e}]$ で増加し、$[\mathrm{e}, \infty)$ で減少するから、$1^{1/1} < 2^{1/2}$ かつ $3^{1/3} > 4^{1/4} > 5^{1/5} > \cdots$ である。よって最大値は $2^{1/2}$ か $3^{1/3}$ のいずれかである。両者を 6 乗すると $2^3 = 8$ と $3^2 = 9$ となるので、$3^{1/3} > 2^{1/2}$ であり、

$$\max_n n^{1/n} = 3^{1/3}.$$

**小問 (ii)**

$q = 1 - p$ とおく。$k$ 人からなる 1 つのグループを考える。独立性（independence）より、グループ内の誰も酵素（enzyme）をもたない確率（probability）は $q^k$ であり、まとめた検体が陰性になるのはその場合に限られる。したがって、このグループが用いる検査の回数は

$$\begin{cases} 1\text{回} & \text{確率} q^k, \\ 1+k\text{回} & \text{確率} 1-q^k, \end{cases}$$

であり、その期待値は $1 + k(1 - q^k)$ である。グループは $r = N/k$ 個あり、期待値は加法的であるから、検査回数の合計の期待値は

$$E = \frac{N}{k}(1 + k(1 - q^k)) = N\left(1 + \frac{1}{k} - (1-p)^k\right).$$

**小問 (iii)**

小問 (ii) より、$E \le N$ となるための必要十分条件は

$$\frac{1}{k} \le (1-p)^k \iff 1 - p \ge k^{-1/k} = \frac{1}{k^{1/k}}$$

である。ここで正の $k$ 乗根をとった（両辺は正であり、$t \mapsto t^{1/k}$ は増加関数である）。

? **考えてみよう** どの $k$ のとき、$p$ を最も大きくできるか。

与えられた $k$ に対して条件は $p \leq 1 - k^{-1/k}$ であり、この上界は $k^{1/k}$ が最大のとき最大となる。小問 (i) より、任意の整数 $k > 1$ に対して $k^{1/k} \leq 3^{1/3}$ であるから、どのような $k$ を選んでも $p \leq 1 - 3^{-1/3}$ でなければならない。逆に、$N$ は 3 の倍数であるから $k = 3$ はグループの大きさとして許され、$k = 3$ のとき条件はちょうど $p \leq 1 - 3^{-1/3}$ となる（等号のとき $E = N$ である）。よって求める最大値は

$$p = 1 - 3^{-1/3} \approx 0.307.$$

これが $\frac{1}{4}$ を超えることを示す。

$$1 - 3^{-1/3} > \tfrac{1}{4} \iff 3^{-1/3} < \tfrac{3}{4} \iff \tfrac{1}{3} < \tfrac{27}{64} \iff 64 < 81,$$

途中の段階では 3 乗した（正の数に対して 3 乗は大小関係を保つ）。最後の主張は正しいので、$1 - 3^{-1/3} > \frac{1}{4}$ である。

### 小問 (iv)

二項展開（binomial expansion）より

$$(1-p)^k = 1 - kp + \binom{k}{2} p^2 - \cdots.$$

$pk$ が小さければ、$\binom{k}{2} p^2 < \frac{1}{2}(pk)^2$ はさらにずっと小さく、それ以降の項も同様であるから、$(1-p)^k \approx 1 - pk$ である。これを小問 (ii) の式に代入すると

$$E \approx N\left(1 + \frac{1}{k} - (1 - pk)\right) = N\left(\frac{1}{k} + pk\right).$$

（実際、ベルヌーイの不等式（Bernoulli's inequality）$(1-p)^k \geq 1 - pk$ により、この近似が $E$ を過小評価することはない。また展開を 1 項先で打ち切れば、誤差が高々 $\frac{1}{2}(pk)^2 N$ であることがわかる。）

$p = 0.01$、$k = 10$ のとき $pk = 0.1$ であり、これは小さい。このとき

$$E \approx N\left(\frac{1}{10} + 0.1\right) = 0.2N,$$

となるので、この方式では、全員を個別に検査する場合の $N$ 回のうち、わずか約 20% の検査で済む。

## 検算

- $p = 0.01$、$k = 10$ で厳密に計算すると、$E/N = 1.1 - 0.99^{10} \approx 1.1 - 0.9044 = 0.1956$、すなわち約 19.6% であり、近似値との差は誤差の上界 $\frac{1}{2}(0.1)^2 = 0.005$ 以内に収まっている。

- $p = 1 - 3^{-1/3}$、$k = 3$ のとき、$(1-p)^3 = \frac{1}{3}$ であるから $E = N(1 + \frac{1}{3} - \frac{1}{3}) = N$ となり、ちょうど境界上にある。
- 小問 (ii) の妥当性：$p \to 0$ のとき $E \to N/k$（各グループ 1 回の検査）、$p \to 1$ のとき $E \to N(1 + \frac{1}{k})$（すべてのグループが再検査される）となる。この公式は、小さなグループについてすべての結果を列挙することにより、コンピュータでも確認した。

## よくある間違い

- 小問 (i) の答えを $e^{1/e}$ としてしまう。$n$ は整数でなければならないので、$2^{1/2}$ と $3^{1/3}$ を比較する必要がある。
- 小問 (iii) で $k = 3$ の場合だけを確かめ、他のグループの大きさではより大きな $p$ が許されない理由を説明しない。
- 陽性のグループで、まとめた検体の検査そのものを数え忘れ、$k+1$ 回ではなく $k$ 回としてしまう。

## STEP のその先へ

これは Dorfman の群検査（group testing）方式であり、第二次世界大戦中に兵士の血液検体をスクリーニングするために考案され、COVID-19 の検査で再び脚光を浴びた。$\frac{1}{k} + pk$ を最小化すると $k \approx 1/\sqrt{p}$（したがってここでは $k = 10$ が最適）となり、検査回数は約 $2\sqrt{p}\,N$ となる。情報理論（information theory）で研究されている、より巧妙な適応的方式を用いれば、さらに効率を上げることができる。背景にある確率論については、Blitzstein and Hwang, *Introduction to Probability*（確率論入門）を参照されたい。

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。