解答例

# STEP II 2000、第 13 問

確率・統計 ・ 捕獲再捕獲法による推定

**STEP II 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-s2-q13

この問題は、*Beyond STEP* の第 13.2 節「ハタネズミを数える」で詳しく解説されている：
https://beyondstep.keithmcnulty.org/ja/ch13/voles.html

## 鍵となる考え方

これは捕獲再捕獲法（capture–recapture）の問題である。2 回目の標本は $N$ 匹のハタネズミから無作為に選んだ 200 匹であり、そのうち 200 匹に印が付いているので、標本中の印付きハタネズミの数は超幾何分布（hypergeometric distribution）に従う。$p_N$ が最大となる場所を求めるには、微分してはいけない（$N$ の階乗（factorial）は、簡単な意味では微分できない）。その代わりに、比 $p_N/p_{N-1}$ を通して隣り合う項を比較する。この比ではほとんどすべてが約分される。最後の恒等式は労せずして得られる。3 回目の標本についての確率の和は 1 にならなければならないからである。

## 解答

### 確率 $p_N$

2 回目の標本を、$N$ 匹のハタネズミから 200 匹を選ぶこととしてモデル化し、大きさ 200 のどの部分集合も同様に確からしいとする。そのような部分集合は $\binom{N}{200}$ 個ある。$N$ 匹のうち 200 匹には印があり、$N-200$ 匹には印がない。印付きのハタネズミをちょうど 11 匹含む標本は、印付きのものから 11 匹、印のないものから 189 匹を選ぶこ

とにあたるので、

$$p_N = \frac{\binom{200}{11}\binom{N-200}{189}}{\binom{N}{200}}.$$

$N$ を含む 2 つの二項係数（binomial coefficient）を階乗で書くと、

$$\begin{aligned} p_N &= \binom{200}{11} \cdot \frac{(N-200)!}{189!\,(N-389)!} \cdot \frac{200!\,(N-200)!}{N!} \\ &= k\,\frac{((N-200)!)^2}{N!\,(N-389)!}, \qquad \text{ただし} k = \binom{200}{11}\frac{200!}{189!}, \end{aligned}$$

であり、$k$ は $N$ によらない。これが求める形である。（ここでは $N \geq 389$ が必要である。2 つの標本を合わせると、$200 + 189 = 389$ 匹の異なるハタネズミが含まれるからである。）

### $p_N$ の最大化

$N \geq 390$ に対して、$p_N$ を $p_{N-1}$ で割る。$(N-200)! = (N-200)\,(N-201)!$、$N! = N\,(N-1)!$、$(N-389)! = (N-389)\,(N-390)!$ を用いると、次の部分を除いてすべて約分される。

$$\frac{p_N}{p_{N-1}} = \frac{(N-200)^2}{N(N-389)}.$$

分子も分母も正なので、

$$\begin{aligned} p_N > p_{N-1} &\iff (N-200)^2 > N(N-389) \\ &\iff N^2 - 400N + 40000 > N^2 - 389N \\ &\iff 40000 > 11N \\ &\iff N < 3636.\overline{36}. \end{aligned}$$

**考えてみよう** このことから、数列 $p_{389}, p_{390}, \ldots$ の形について何がわかるか？

$390 \leq N \leq 3636$ では比は 1 より大きいので、数列は $p_{3636}$ まで狭義単調に増加する。$N \geq 3637$ では比は 1 より小さい（等号が成り立つには $11N = 40000$ でなければならないが、これは整数解をもたない）ので、数列はそこから先は狭義単調に減少する。よって $p_N$ はただ 1 つの最大値をとり、それは

$$N = 3636.$$

のときである。これは $\frac{200\times200}{11}$ の整数部分である。標本中の印付きの割合 $\frac{11}{200}$ が、母集団中の印付きの割合 $\frac{200}{N}$ に一致するはずだと考えれば、こう推測できるであろう。

### 3回目の標本

母集団の大きさを、いま求めた推定値 3636 とする。2 回目の放獣の後、印付きのハタネズミは最初の 200 匹と新たに印を付けた 189 匹なので、389 匹が印付き、$3636-389=3247$ 匹が印なしである。先ほどとまったく同様に考えると、大きさ 200 の無作為標本がちょうど $j$ 匹の印付きハタネズミを含む確率は

$$\frac{\binom{389}{j}\binom{3247}{200-j}}{\binom{3636}{200}}, \qquad j=0,1,\ldots,200.$$

である。$389 \geq 200$ かつ $3247 \geq 200$ なので、0 から 200 までのどの $j$ も起こりうり、これらで起こりうる結果がすべて尽くされる。これらは互いに排反なので、確率の和は 1 である。

$$\sum_{j=0}^{200} \frac{\binom{389}{j}\binom{3247}{200-j}}{\binom{3636}{200}} = 1.$$

両辺に $\binom{3636}{200}$ を掛けると

$$\sum_{j=0}^{200} \binom{389}{j}\binom{3247}{200-j} = \binom{3636}{200},$$

となり、示すべき式が得られた。

## 検算

- 整数の二項係数を用いて $p_N$ を厳密に計算すると、$N$ のいくつかの値（たとえば 389、1000、5000）で階乗の公式と一致する。また、$389 \leq N < 20000$ の範囲で探索すると、最大値は $N=3636$ でとられる。
- 転換点では $p_{3636}/p_{3635} = 3436^2/(3636\times 3246) \approx 1.0003 > 1$ であるが、$p_{3637}/p_{3636} < 1$ である。
- 最後の恒等式の両辺を整数として厳密に計算すると、等しくなる。また、当然ながら $389+3247=3636$ であることに注意せよ。

## よくある間違い

- $N$ を連続変数とみなして微分してしまうこと。代わりに隣り合う項の比を用い、不等式を厳密に判定すること。

- 2 回目の放獣の後に印が付いているのは 200 匹ではなく 389 匹であることを忘れること。
- 「$N = 3636.36$ で最大」と答えること。母集団の大きさは整数でなければならず、隣り合うどちらの整数で大きくなるかを確かめるべきである。

## STEP のその先へ

観測されたデータが最も起こりやすくなるようにパラメータの値を選ぶ方法は**最尤法**（maximum likelihood）とよばれ、大学 1 年次の統計学の中心的な考え方である。ここでは、捕獲再捕獲法に対するリンカーン–ピーターセン推定量（Lincoln–Petersen estimator）が得られる。最後の恒等式はヴァンデルモンドの恒等式（Vandermonde's identity）であり、ここで用いた論法、すなわち同じものを 2 通りに数える方法は、組合せ論における強力な手法である。Grimmett and Welsh, *Probability: An Introduction*（確率論入門）では、超幾何分布とこの種の論法が扱われている。

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。