13.2 ハタネズミを数える

ワークシート

問題

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

ヒント

まずは自力で問題に取り組んでみよう。行き詰まったら、ヒントを**一つずつ**読み、そのたびに問題に戻ること。**手を止めて、やってみよう**と記したヒントが重要な段階である。各ヒントの後には計算用の余白がある。

> **考えてみよう**　個体群には N 匹のハタネズミがいて、そのうち 200 匹に印がついている。無作為に 200 匹の標本を取り出す。ちょうど 11 匹に印がついている確率はいくらか。

段階 1　尤度（likelihood）

扱う内容：確率 p_N

> **考えてみよう**　同様に確からしい 200 匹の標本（sample）全体のうち、印のついたハタネズミから 11 匹、印のないハタネズミから 189 匹を選ぶ場合を考えよ。そのうえで、N を含む二項係数（binomial coefficient）を階乗（factorial）を用いて書き下せ。

段階 2 整数の範囲での最大化（maximising）

扱う内容：p_N の最大化

> **手を止めて、やってみよう** N は整数であるから、p_{N+1} と p_N（あるいは p_N と p_{N-1}）を比較せよ。階乗の大部分は約分される。この比（ratio）が 1 より大きくなるのはどのようなときか。

段階 3 3 回目の標本

扱う内容：3 回目の標本

> **考えてみよう** 2 回目の標本を取ったあと、印のついたハタネズミは何匹いるか。N の推定値を用いると、印のないハタネズミは何匹いるか。200 匹からなる 3 回目の標本に、印のついたハタネズミがちょうど j 匹含まれる確率を書け。これらの確率の和はいくつになるか。

背後にある数学

最尤法

最尤法の原理は、1910 年代から 1920 年代にかけてフィッシャー（Ronald Fisher）が発展させたもので、「未知のパラメータを、観測されたデータが最も起こりやすくなる値によって推定せよ」というものである。これは群を抜いて広く使われている推定法であり、データへの曲線のあてはめから、機械学習（machine learning）におけるニューラルネットワーク（neural networks）の訓練にまで及ぶ。後者の「損失関数」（loss functions）は、たいてい対数尤度にマイナスをつけたものである。フィッシャーは、モデルが性質のよいもの（日常的なモデルの大半で成り立つ条件を満たすもの）であれば、標本が大きいとき最尤推定値は近似的に正規分布（normal distribution）に従い、およそ妥当な推定量（estimator）の中で最も精度が高いことを示した。ただし重要な例外もある。次節（Section 13.3）では、未知数がとりうる最大の値、すなわち分布の端点であり、そこでは最尤推定値の振る舞いはまったく異なる。その誤差は近似的に正規分布に従わず、$\frac{1}{\sqrt{n}}$ ではなく $\frac{1}{n}$ のように小さくなっていくのである。

推定はどれほど精確か

図 13.2.1（オンライン版の節全体を参照）は尤度の山が幅広いことを示している。11 匹の再捕獲はそれほど多くの情報ではないのである。大まかに計算してみよう。2 回

目の標本に含まれる印のついたハタネズミの数は、近似的に平均 $\frac{200\times200}{N}$ のポアソン分布（Poisson distribution）に従うので、その標準偏差（standard deviation）はおよそ $\sqrt{11}\approx3.3$ であり、相対誤差は約 30% となる。これはそのまま N の推定値にも引き継がれる。N を 10% 以内の精度で推定するには、生物学者はおよそ 100 匹の再捕獲を必要とするだろう。統計学者はまた、単純な推定値のわずかな偏り（bias）を補正するため、**チャップマン推定量**（Chapman estimator）$\frac{(M+1)(C+1)}{R+1}-1$ を用いる。ここで M と C は 2 回の捕獲の大きさ、R は再捕獲された数である。

ラプラス、フランスを数える

ラプラス（Laplace）は 1802 年、同じ考え方を用いて、国勢調査を行わずにフランスの人口を推定した。彼は 1 年間に国全体で登録された出生数を知っており、さらにいくつかのコミューンを標本として、その人口と出生数の両方を把握していた。人口と出生数の比がどこでも同じであると仮定して、彼はフランスの人口をおよそ 28 百万人と推定し、さらに不確かさの尺度まで与えた。この方法は、動物の個体数について改めて発見された。1890 年代にデンマークのフィヨルドでツノガレイに標識をつけた C. G. J. Petersen と、1930 年代に鳥の足環を用いた Frederick Lincoln によってである。そのため、今でもしばしば**リンカーン–ピーターセン推定量**（Lincoln–Petersen estimator）と呼ばれている。

次の一歩

1. **一般の公式。** M 匹の動物に印をつけ、2 回目の大きさ C の標本に印のついたものが R 匹含まれていたとする。N の最尤推定値は $\frac{MC}{R}$ の整数部分であることを示せ。

2. **ヴァンデルモンド。** a 人の女性と b 人の男性から c 人の委員会を選ぶ方法を数えることにより、$\sum_j \binom{a}{j}\binom{b}{c-j}=\binom{a+b}{c}$ を証明せよ。

3. **硬貨投げ。** 表の出る確率 p が未知の硬貨を n 回投げたところ、表が h 回出た。p の最尤推定値は $\frac{h}{n}$ であることを示せ。

4. **再捕獲ゼロ。** 2 回目の標本に印のついたハタネズミが 1 匹も含まれていなかった場合、最尤推定値はどうなるか。そのことはこの方法について何を教えてくれるか。

さらに学ぶために

- **David Spiegelhalter, *The Art of Statistics***（Pelican）。推定と不確かさについて。
- **George Seber, *The Estimation of Animal Abundance***（Blackburn）。標識再捕獲法につ

いての標準的な参考書。

- **John Rice, *Mathematical Statistics and Data Analysis***（Duxbury）。最尤法を含む、数理統計学の入門書。

詳しい解説付きの解答は第 13.2 節にある：

https://beyondstep.keithmcnulty.org/ja/ch13/voles.html

毎週 1 問の STEP の問題を丁寧に解き明かす The STEP Up。メールで無料配信中：keithmcnulty.substack.com