13.3 ドイツ戦車問題

ワークシート

問題

STEP III 1998、問 14

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/98-s3-q14

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　M を $[0, N]$ 上の独立な n 個の一様分布に従う値の最大値とするとき、$\mathrm{P}(M \leq m)$ はいくらか。

段階 1 どちらも不偏である

扱う内容：Z_1 と Z_2 の平均

考えてみよう Z_2 の平均はすぐに求まる。各 X_i の平均はいくらか。

手を止めて、やってみよう Z_1 については、最大値 M の分布を求めよ。$M \leq m$ となるのはどのようなときか。それを微分して確率密度関数（probability density function）を求め、そこから $\mathrm{E}(M)$ を求めよ。

段階 2　それぞれの分散

扱う内容：分散

考えてみよう　$\mathrm{E}(M^2)$ を求め、それにより $\mathrm{Var}(M)$ を求めよ。次に定数倍して $\mathrm{Var}(Z_1)$ を得よ。Z_2 については、一つの X_i の分散はいくらか。X_i は互いに独立（independent）である。

段階 3 どちらがよいか

扱う内容：どちらを選ぶべきか

> **考えてみよう** 一方の分散を他方の分散で割ってみよ。n が大きくなるとき、その比はどのように振る舞うか。また、どちらかの推定量（estimator）が、実際に観測したシリアル番号より小さくなることはあり得るか。

背後にある数学

十分性

なぜ最大値はこれほどよいのだろうか。最大値 M が与えられると、残りの値は N が何であっても単に $[0, M]$ 上の一様分布に従う。つまり、それらは N についてそれ以上の情報をもたない。最大値は**十分統計量**（sufficient statistic）なのである。ラオ（C. R. Rao）とブラックウェル（David Blackwell）による定理（1945–47 年）は、どんな不偏推定量も、十分統計量を与えたときの条件付き期待値に置き換えることで改良できる（少なくとも悪くはならない）ことを述べており、ここではこの操作によって Z_2 が Z_1 に変わる。さらに**レーマン–シェフェの定理**（Lehmann–Scheffé theorem）により、Z_1 が分散の最も小さい不偏推定量であることがわかる。また、Z_1 の分散が $\frac{1}{n^2}$ のように減少し、多くの推定量の $\frac{1}{n}$ より速いことにも注意しよう。これは N が分布の端点であり、データがそこに押し寄せるように集まるからである。

離散版

実際のシリアル番号は整数 $1, 2, \ldots, N$ であり、非復元抽出（sampling without replacement）で得られる。このとき最良の不偏推定値は $\hat{N} = m + \frac{m}{k} - 1$ である。ここで m は観測された k 個のシリアル番号の最大値であり、この式は最大値に観測値どうしの平均的な間隔を加えたものになっている。主にギアボックスと車輪のシリアル番号に基づくこの種の推定を用いて、連合国の統計学者たちはドイツの生産数を月に約 246

両と推定した。これに対し、のちにドイツ側の記録で判明した数は 245 両、情報部の報告が示唆していた数は 1400 両であった。

歴史の中のシリアル番号

この方法はその後も何度も用いられてきた。たとえば、新しい消費者向け製品が何台作られたかを、最初に売られた数台のシリアル番号から分析者が推定した例がある。メーカーもこの教訓を学び、今では多くのメーカーが、単純な連番として読み取れないようにシリアル番号を符号化している。Richard Ruggles と Henry Brodie は、1947 年に *Journal of the American Statistical Association* で戦時中のこの研究について記している。

次の一歩

1. **平均二乗誤差。** 偏りのある推定量 M について、$\mathrm{E}(M-N)^2 = \mathrm{Var}(M) + \big(\mathrm{E}(M)-N\big)^2$ が成り立つ。これを計算し、$\mathrm{Var}(Z_1)$ と比較せよ。また、$\mathrm{E}(cM-N)^2$ を最小にする定数 c を求めよ。

2. **最小値も。** $(n+1)\min\{X_i\}$ もまた N の不偏推定量であることを示し、その分散を求めよ。これはよい推定量だろうか。

3. **離散版の公式。** シリアル番号 $1,\dots,N$ から k 個を非復元抽出し、その最大値を m とする。$\mathrm{P}(\max \le m) = \binom{m}{k}/\binom{N}{k}$ を示し、$k=1$ のとき $m + \frac{m}{k} - 1$ が不偏であることを確かめよ。

4. **戦車。** 鹵獲した 5 両の戦車のシリアル番号が 19、40、42、60、98 であるとき、製造された戦車の数を推定せよ。

さらに学ぶために

- **Richard Ruggles & Henry Brodie, “An empirical approach to economic intelligence in World War II”**, *Journal of the American Statistical Association* 42 (1947)。当事者による原論文。
- **John Rice, *Mathematical Statistics and Data Analysis***（Duxbury）。偏り、分散、十分性について。
- **David Spiegelhalter, *The Art of Statistics***（Pelican）。

詳しい解説付きの解答は第 13.3 節にある：

https://beyondstep.keithmcnulty.org/ja/ch13/german-tanks.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