解答例

STEP II 2001、第 14 問

確率・統計 ・ 最小分散推定量

> **STEP II 2001、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/01-s2-q14

鍵となる考え方

T は独立な 2 つのベルヌーイ変数（Bernoulli variable）の重み付き結合であるから、その平均と分散は、期待値の線形性（linearity of expectation）と、独立な X, Y に対する法則 $\operatorname{Var}(aX + cY) = a^2 \operatorname{Var}(X) + c^2 \operatorname{Var}(Y)$ からただちに求まる。重みは、λ が何であっても平均が p になるように選ばれている。すなわち T は p の**不偏推定量**（unbiased estimator）であり、この問題はその中で最良のものを求めているのである。最後の部分では、中心極限定理（central limit theorem）によって $\overline{T}$ は近似的に正規分布に従い、平均が一定の正規変数は分散が最小のときに最も集中する。

解答

T の平均と分散

ここで X と Y はベルヌーイ変数であり、

$$\mathrm{E}(X) = p, \quad \operatorname{Var}(X) = p(1-p), \qquad \mathrm{E}(Y) = 2p, \quad \operatorname{Var}(Y) = 2p(1-2p).$$

である。（確率 q で値 1 をとり、それ以外では 0 をとる変数については、$\mathrm{E} = q$ かつ $\mathrm{E}(\cdot^2) = q$ であるから、分散は $q - q^2 = q(1-q)$ である。）期待値の線形性により、

$$\mathrm{E}(T) = \lambda p + \tfrac{1}{2}(1-\lambda)\cdot 2p = \lambda p + (1-\lambda)p = p,$$

となり、求める結果が得られる。硬貨は独立であるから X と Y は独立であり、したがって共分散（covariance）の項は現れず、

$$\begin{aligned}\operatorname{Var}(T) &= \lambda^2 \operatorname{Var}(X) + \tfrac{1}{4}(1-\lambda)^2 \operatorname{Var}(Y) \\ &= \lambda^2 p(1-p) + \tfrac{1}{2}(1-\lambda)^2 p(1-2p).\end{aligned}$$

となる。

分散の最小化

$V(\lambda) = \mathrm{Var}(T)$ とおく。これは λ の 2 次式で、最高次の係数 $p(1-p) + \frac{1}{2}p(1-2p)$ は正である（$0 < p < \frac{1}{2}$ より両項とも正）。したがって最小値はただ 1 つで、それは $V'(\lambda) = 0$ となるところでとられる：

$$V'(\lambda) = 2\lambda p(1-p) - (1-\lambda)p(1-2p) = 0.$$

$p > 0$ で割り、λ を含む項をまとめると、

$$\lambda[2(1-p) + (1-2p)] = 1 - 2p \quad \Longrightarrow \quad \lambda = \frac{1-2p}{3-4p},$$

となり、これは問題文で与えられた値である。（ここで $3 - 4p > 1 > 0$ であるから、0 で割ることにはならない。）

最小値そのものも後で必要になる。$a = p(1-p)$, $c = \frac{1}{2}p(1-2p)$ とおくと $V = a\lambda^2 + c(1-\lambda)^2$ であり、これは $\lambda = c/(a+c)$（上の値に一致する）で最小となって、

$$V_{\min} = a\Big(\frac{c}{a+c}\Big)^2 + c\Big(\frac{a}{a+c}\Big)^2 = \frac{ac(a+c)}{(a+c)^2} = \frac{ac}{a+c}.$$

を得る。ここで $a + c = \frac{1}{2}p[2(1-p) + (1-2p)] = \frac{1}{2}p(3-4p)$, $ac = \frac{1}{2}p^2(1-p)(1-2p)$ であるから、

$$V_{\min} = \frac{p(1-p)(1-2p)}{3-4p}.$$

である。

n 回の投げの平均

n 回の投げにおける T の値を $T_1, \ldots, T_n$ とする。これらは独立で、いずれも平均 p、分散 $V(\lambda)$ をもつ。よって

$$\mathrm{E}(\overline{T}) = p, \qquad \mathrm{Var}(\overline{T}) = \frac{V(\lambda)}{n}.$$

である。

> **考えてみよう** T は 4 つの値しかとらない。なぜ $\overline{T}$ を正規分布に従うものとして扱ってよいのか。

$\overline{T}$ は有限の分散をもつ独立同分布（independent, identically distributed）な n 個の変数の平均であり、しかも $n > 30$ は十分大きいので、中心極限定理により、1 つの

T の分布の形がどうであれ、$\overline{T}$ は近似的に $\mathrm{N}(p,\, V(\lambda)/n)$ に従う。$\sigma_\lambda = \sqrt{V(\lambda)/n}$, $Z = (\overline{T} - p)/\sigma_\lambda \approx \mathrm{N}(0,1)$ とおくと、

$$\mathrm{P}(|\overline{T} - p| < b) = \mathrm{P}\Big(|Z| < \frac{b}{\sigma_\lambda}\Big) \approx \Phi\Big(\frac{b}{\sigma_\lambda}\Big) - \Phi\Big(-\frac{b}{\sigma_\lambda}\Big) = 2\Phi\Big(\frac{b}{\sigma_\lambda}\Big) - 1,$$

となる。ここで対称性 $\Phi(-z) = 1 - \Phi(z)$ を用いた。

$b > 0$ は固定されており Φ は狭義単調増加であるから、この値は b/σ_λ が最大のとき、すなわち σ_λ が最小のときに最大となり、それは $V(\lambda)$ が最小となるときである。それが起こるのは $\lambda = (1-2p)/(3-4p)$ のときで、このとき

$$\sigma_\lambda^2 = \frac{V_{\min}}{n} = \frac{p(1-p)(1-2p)}{(3-4p)n} = s^2.$$

である。したがって、λ を動かしたときの $\mathrm{P}(|\overline{T} - p| < b)$ の最大値は近似的に $2\Phi(b/s) - 1$ であり、求める結果が得られた。

検算

- $\lambda = 1$ のとき $T = X$ であり、公式は $\mathrm{Var}(T) = p(1-p)$ を与える。$\lambda = 0$ のとき $T = \frac{1}{2}Y$ であり、公式は $\frac{1}{4} \cdot 2p(1-2p)$ を与える。どちらも正しい。
- $V_{\min}$ はこのどちらよりも小さい。実際、それは $p(1-p)$ の $\frac{1-2p}{3-4p} < 1$ 倍である。2 枚の硬貨を組み合わせるほうが、どちらか一方だけを使うよりも優れている。
- 数値で確かめる。$p = 0.3$, $n = 100$, $b = 0.03$ とすると、最良の λ は $\frac{2}{9}$、$s \approx 0.0216$ であり、$2\Phi(b/s) - 1 \approx 0.835$ となる。100 回の投げを 40 000 回繰り返すシミュレーションでは 0.836 が得られ、一方 $\lambda = 1$（硬貨 A のみ）ではおよそ 0.487 にとどまった。

よくある間違い

- $\mathrm{Var}(\frac{1}{2}(1-\lambda)Y) = \frac{1}{2}(1-\lambda)\,\mathrm{Var}(Y)$ と書いてしまうこと。定数は 2 乗されて分散の外に出る。
- 分散を最小化するとなぜ確率が最大になるのか、その**理由**を述べ忘れること。それには Φ が増加関数であることと、すべての λ に対して $\overline{T}$ の平均が p であることが必要である。

STEP のその先へ

これは推定理論（estimation theory）への最初の入り口である。p のすべての不偏推定量 $\lambda X + \frac{1}{2}(1-\lambda)Y$ の中から、分散が最小のものを見つけたのである。答えは各観測値にその分散に反比例する重みを与えるもので、これは統計学において精度の異なる測定値を組み合わせる際に使われる規則（逆分散重み付け、inverse-variance weighting）と同じである。また最小値 $ac/(a+c)$ は、「並列抵抗」の公式 $1/V_{\min} = 1/a + 1/c$ になっている。最後のステップの背後にある中心極限定理については、Grimmett and Welsh, *Probability: An Introduction*（確率論入門）が扱っている。

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。