解答例

STEP II 1996、第 14 問

確率・統計 ・ 積率と中心極限定理

> **STEP II 1996、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/96-s2-q14

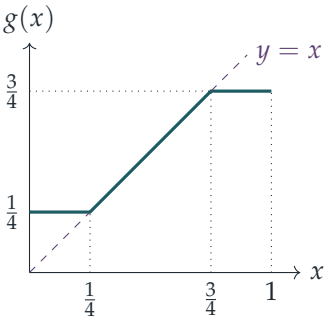
鍵となる考え方

Y は X の両端を「切り詰めた」ものである。すなわち、$\frac{1}{4}$ より下にある確率はすべて一点 $\frac{1}{4}$ に積み上げられ、$\frac{3}{4}$ より上にある確率はすべて $\frac{3}{4}$ に積み上げられる。したがって Y は、二つの点質量（point mass）と連続な中央部分をもつ混合型確率変数（mixed random variable）であり、任意の期待値（expectation）$\mathrm{E}(g(Y))$ も同じように三つの部分に分かれる。最後の小問では非常に多くの独立なコピーの和が現れるので、中心極限定理（central limit theorem）を用いることになる。そのために必要なのは Y の平均と分散（variance）だけであり、それはそれまでの小問ですでに求めてある。

解答

Y のモーメント

$X \sim \mathrm{U}[0,1]$ であるから、$\mathrm{P}(X \leq \frac{1}{4}) = \mathrm{P}(X \geq \frac{3}{4}) = \frac{1}{4}$ であり、$[\frac{1}{4}, \frac{3}{4}]$ 上で X の確率密度（density）は 1 である。下に図示した切り詰め関数 g を用いて $Y = g(X)$ と書けば、$\mathrm{E}(Y^n) = \int_0^1 g(x)^n \,\mathrm{d}x$ を得る。



積分を $\frac{1}{4}$ と $\frac{3}{4}$ で分割すると、

$$\begin{aligned}\mathrm{E}(Y^n) &= \left(\tfrac{1}{4}\right)^n \cdot \tfrac{1}{4} + \int_{1/4}^{3/4} x^n \,\mathrm{d}x + \left(\tfrac{3}{4}\right)^n \cdot \tfrac{1}{4} \\ &= \frac{1}{4^{n+1}} + \frac{3^{n+1}-1}{(n+1)4^{n+1}} + \frac{3^n}{4^{n+1}} \\ &= \frac{(n+1)(1+3^n)+3^{n+1}-1}{(n+1)4^{n+1}}.\end{aligned}$$

分子は $n+1+(n+1)3^n+3\cdot 3^n-1 = n+(n+4)3^n$ であるから、

$$\boxed{\mathrm{E}(Y^n) = \frac{n+(n+4)\,3^n}{(n+1)\,4^{n+1}}} \qquad (n \geq 1).$$

X との比較

一様分布に従う確率変数については $\mathrm{E}(X^n) = \int_0^1 x^n \,\mathrm{d}x = \dfrac{1}{n+1}$ である。

$n=1$ とすると、公式より $\mathrm{E}(Y) = \dfrac{1+5\cdot 3}{2\cdot 16} = \dfrac{1}{2} = \mathrm{E}(X)$ となり、題意を満たす。（これは驚くことではない。切り詰めによって両端の質量は $\frac{1}{2}$ に関して対称に、同じ量だけ内側へ移動するからである。）

$n=2$ とすると $\mathrm{E}(Y^2) = \dfrac{2+6\cdot 9}{3\cdot 64} = \dfrac{56}{192} = \dfrac{7}{24}$ であり、一方 $\mathrm{E}(X^2) = \frac{1}{3} = \frac{8}{24}$ である。よって

$$\mathrm{E}(X^2) - \mathrm{E}(Y^2) = \frac{8}{24} - \frac{7}{24} = \frac{1}{24},$$

となり、題意を満たす。

一般の $n \geq 2$ については、二つのモーメント（moment）を共通の分母 $(n+1)4^{n+1}$ で表す：

$$\mathrm{E}(X^n) = \frac{4^{n+1}}{(n+1)4^{n+1}}.$$

したがって $4^{n+1} > n + (n+4)3^n$ を示せばよい。$4^{n+1} = 4(3+1)^n$ と二項定理（binomial theorem）を用い、上位の二項を別にしておくと、

$$4^{n+1} = 4 \cdot 3^n + 4n \cdot 3^{n-1} + 4 \sum_{k=0}^{n-2} \binom{n}{k} 3^k.$$

ここから $n + (n+4)3^n = 4 \cdot 3^n + 3n \cdot 3^{n-1} + n$ を引くと、

$$4^{n+1} - (n + (n+4)3^n) = n\,3^{n-1} - n + 4 \sum_{k=0}^{n-2} \binom{n}{k} 3^k.$$

考えてみよう この差はなぜ正なのか。

$n \geq 2$ のとき $3^{n-1} \geq 3$ であるから、$n\,3^{n-1} - n \geq 2n > 0$ である。和の部分も正である。1 に等しい $k = 0$ の項を含み、どの項も非負だからである。よって差は真に正であり、

$$\mathrm{E}(X^n) > \mathrm{E}(Y^n) \qquad \text{任意の} n \geq 2.$$

が成り立つ。（$n = 1$ のとき差は $3^0 - 1 + 0 = 0$ であり、$\mathrm{E}(X) = \mathrm{E}(Y)$ と一致する。）

240 000 個のコピーの和

前の小問より、

$$\mathrm{Var}(Y) = \mathrm{E}(Y^2) - (\mathrm{E}(Y))^2 = \frac{7}{24} - \frac{1}{4} = \frac{1}{24}.$$

$N = 240\,000$ として $S = Y_1 + \cdots + Y_N$ とおく。Y_i は独立であるから、期待値も分散も足し合わせることができる：

$$\mathrm{E}(S) = \frac{N}{2} = 120\,000, \qquad \mathrm{Var}(S) = \frac{N}{24} = 10\,000,$$

よって S の標準偏差（standard deviation）は 100 である。Y_i は有限の分散をもつ独立同分布（independent and identically distributed）の確率変数であり、N は非常に大きいので、中心極限定理により S は近似的に $\mathrm{N}(120\,000, 100^2)$ に従う。よって

$$\mathrm{P}(S < K) \approx \Phi\Big(\frac{K - 120\,000}{100}\Big) = \frac{3}{4} \quad \Longrightarrow \quad \frac{K - 120\,000}{100} = \Phi^{-1}(0.75) \approx 0.6745,$$

したがって

$$K \approx 120\,000 + 67.45 \approx 120\,067.$$

連続性補正（continuity correction）は必要ない。Y は整数値をとる確率変数ではないし、そもそも標準偏差が 100 であれば、そのような補正はこの精度では結果に何の影響も与えない。

検算

- $\frac{1}{4} \le Y \le \frac{3}{4}$ であるから、$\mathrm{E}(Y^n)$ の公式の値は $(\frac{1}{4})^n$ と $(\frac{3}{4})^n$ の間になければならないが、実際そうなっている。$n = 3$ のとき公式は $\frac{3+7\cdot 27}{4\cdot 256} = \frac{192}{1024} = \frac{3}{16} = 0.1875$ を与え、$g(x)^3$ を中点則（midpoint rule）で数値積分した値と小数第 10 位まで一致する。
- $\mathrm{Var}(Y) = \frac{1}{24} < \frac{1}{12} = \mathrm{Var}(X)$ である。切り詰めは値を平均の方へ引き寄せるので、散らばりは小さくなるはずである。
- Y の 2400 個のコピーの和を 20 000 回シミュレーションしたところ、標準化した上側四分位点（upper quartile）は約 0.678 となり、0.6745 に近かった。これは正規近似（normal approximation）を裏づけている。

よくある間違い

- Y を $[\frac{1}{4}, \frac{3}{4}]$ 上の一様分布とみなしてしまい、両端にある大きさ $\frac{1}{4}$ の点質量を忘れること。この区間上の一様分布に従う確率変数の密度は 1 ではなく 2 である。
- $N\,\mathrm{Var}(Y)$ ではなく $\mathrm{Var}(S) = N^2\,\mathrm{Var}(Y)$ を用いてしまうこと。独立性により分散は足し合わされるので、標準偏差は $\sqrt{N}$ に比例して増える。
- 最後の段階で符号を間違えること。$\mathrm{P}(S < K) = \frac{3}{4} > \frac{1}{2}$ であるから、K は平均より大きくなければならない。

STEP のその先へ

Y は X を切り詰めた（あるいは「ウィンザー化（Winsorised）」した）ものであり、その分布は離散でも連続でもない。測度論的確率論（measure-theoretic probability）では、このような分布も分布関数（distribution function）を通じて統一的に扱われ、あらゆる期待値は単一の積分 $\int g\,\mathrm{d}F$ となる。平均が等しいもとでの不等式 $\mathrm{E}(X^n) > \mathrm{E}(Y^n)$ は凸順序（convex ordering）の一例である。X は Y の「平均保存拡散（mean-preserving spread）」であり、したがってすべての凸（convex）関数 ϕ に対して $\mathrm{E}(\phi(X)) \ge \mathrm{E}(\phi(Y))$ が成り立つ。最後に用いた中心極限定理の証明は、Grimmett と Welsh の *Probability: An Introduction*（確率論入門）に載っている。

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。