解答例

STEP I 1997、第 12 問

確率・統計 ・ 中心極限定理

STEP I 1997、問 12

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/97-s1-q12

鍵となる考え方

二次方程式（quadratic equation）を具体的に解き、X を T の簡単な関数として表す。そうすれば X に関する問いはすべて T に関する問いに置き換わり、T の分布は完全にわかっている。うまい簡略化として、$1-T$ もまた $[0,1]$ 上の一様分布（uniform distribution）に従うことが使える。最後の小問は多数の独立（independent）な複製の和に関するものであり、これは中心極限定理（central limit theorem）を使えという合図である。

解答

X の分布

解の公式（quadratic formula）より、$x^2+2x+T=0$ の解は $x=-1\pm\sqrt{1-T}$ である。$0\le T\le 1$ であるからこれらは実数であり、大きい方の解は正の符号をとったものである：

$$X=-1+\sqrt{1-T}.$$

T が $[0,1]$ を動くとき、X は $[-1,0]$ を動く。

$U=1-T$ とおく。T は $[0,1]$ 上の一様分布に従うので、U もそうである。区間をその中点に関して折り返しても、一様な密度は変わらないからである。よって $U\sim \mathrm{U}[0,1]$ として $X=\sqrt{U}-1$ である。

$X > -\frac{1}{3}$ となる確率

$X+1 = \sqrt{U}$ であり、両辺とも非負であるから、

$$X > -\tfrac{1}{3} \iff \sqrt{U} > \tfrac{2}{3} \iff U > \tfrac{4}{9}.$$

両辺が非負なので、2 乗しても不等号の向きは変わらない。したがって

$$\mathrm{P}\left(X > -\tfrac{1}{3}\right) = \mathrm{P}\left(U > \tfrac{4}{9}\right) = 1 - \tfrac{4}{9} = \frac{5}{9}.$$

（同値な言い方をすれば、$T < \frac{5}{9}$ である。）

平均と分散

$U \sim \mathrm{U}[0,1]$ の密度は $[0,1]$ 上で 1 であるから、

$$\mathrm{E}(\sqrt{U}) = \int_0^1 u^{1/2}\,\mathrm{d}u = \frac{2}{3}, \qquad \mathrm{E}((\sqrt{U})^2) = \mathrm{E}(U) = \frac{1}{2}.$$

よって

$$\mathrm{E}(X) = \mathrm{E}(\sqrt{U}) - 1 = \frac{2}{3} - 1 = -\frac{1}{3}.$$

定数を引いても分散（variance）は変わらないので、

$$\mathrm{Var}(X) = \mathrm{Var}(\sqrt{U}) = \mathrm{E}(U) - (\mathrm{E}\sqrt{U})^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18},$$

となり、示すべきことが示された。

> **考えてみよう** これらのモーメントを求めるのに、X の確率密度関数（probability density function）は必要だろうか？

必要ない。任意の関数 g について $\mathrm{E}(g(U)) = \int_0^1 g(u)\,\mathrm{d}u$ であるから、U を直接扱えばよい。それでも密度を知りたければ、$-1 \le x \le 0$ に対して $\mathrm{P}(X \le x) = \mathrm{P}(U \le (x+1)^2) = (x+1)^2$ であるから、X は $[-1,0]$ 上で密度 $2(x+1)$ をもつ。$x \cdot 2(x+1)$ をこの区間で積分しても $-\frac{1}{3}$ が得られる。

800 個の解の和

各試行が独立であるから X_i も独立であり、それぞれ平均 $-\frac{1}{3}$、分散 $\frac{1}{18}$ をもつ。よって

$$\mathrm{E}(Y) = 800 \cdot \left(-\tfrac{1}{3}\right) = -\frac{800}{3}, \qquad \mathrm{Var}(Y) = 800 \cdot \frac{1}{18} = \frac{400}{9},$$

ここで分散を足し合わせる際に独立性を用いている。Y の標準偏差（standard deviation）は $\frac{20}{3}$ である。

800 は大きいので、中心極限定理により Y は近似的に $\mathrm{N}\left(-\frac{800}{3}, \frac{400}{9}\right)$ に従う。求めたいのは

$$\mathrm{P}(Y \le K) \approx \Phi\left(\frac{K+\frac{800}{3}}{20/3}\right) = 0.08.$$

となる K である。数表より $\Phi(1.405) \approx 0.92$ であるから、対称性により $\Phi(-1.405) \approx 0.08$ である。したがって

$$\frac{K+\frac{800}{3}}{20/3} \approx -1.405,$$

$$K \approx -\frac{800}{3} - 1.405 \times \frac{20}{3} \approx -266.67 - 9.37 \approx -276.$$

考えてみよう 連続性補正（continuity correction）は必要ないのだろうか？

必要ない。各 X_i は連続型確率変数であるから Y もそうであり、連続性補正は不要である。この補正は、二項分布（binomial distribution）のような離散分布を正規分布で近似するときにのみ必要となる。

検算

- 解は正しい。$T = 0$ のとき方程式は $x(x+2) = 0$ となり、その大きい方の解は $0 = -1 + \sqrt{1}$ である。
- $\mathrm{E}(X) = -\frac{1}{3}$ は $[-1, 0]$ に含まれる。また密度 $2(x+1)$ は 0 の側に偏っているので、平均が中点 $-\frac{1}{2}$ より大きいことは納得がいく。
- T の値を 400 000 個用いたシミュレーションでは、$\mathrm{P}(X > -\frac{1}{3}) \approx 0.555$、平均 ≈ -0.3337、分散 ≈ 0.0556 が得られた。Y を直接 4000 回シミュレーションすると $\mathrm{P}(Y \le -276.03) \approx 0.083$ となり、0.08 と整合する。

よくある間違い

- 解の公式で負の符号をとってしまう。それでは小さい方の解になる。
- $\mathrm{Var}(Y) = 800^2\,\mathrm{Var}(X)$ と書いてしまう。これは $800X_1$ の分散であって、800 個の独立な複製の和の分散ではない。
- -1.405 の代わりに $+1.405$ を使ってしまう。確率 0.08 は下側の裾にあるので、K は平均より小さくなければならない。

STEP のその先へ

一様分布に従う U を用いて $X = g(U)$ と表す手法は、逆向きにも使える。F が任意の連続な分布関数（distribution function）ならば、$F^{-1}(U)$ の分布関数は F である。この**逆変換法**（inverse transform method）は、コンピュータが一様乱数から確率変数をシミュレーションする方法である。ここでブラックボックスとして用いた中心極限定理は、大学の最初の確率論の講義で、積率母関数（moment generating function）や特性関数（characteristic function）を用いて証明される。Grimmett と Welsh の *Probability: An Introduction*（確率論入門）は、これら両方の考え方を扱っている。

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。