解答例

STEP I 見本問題、第 16 問

確率・統計 ・ 連続型確率変数

STEP I 見本問題、問 16

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/spec-s1-q16

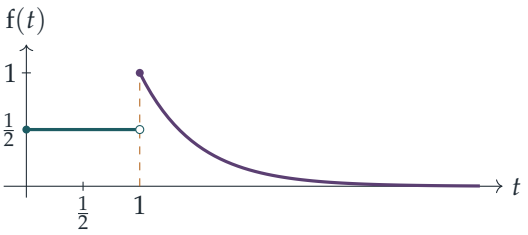
鍵となる考え方

未知数 k は全確率が 1 であることから決まり、それ以外はすべて確率密度（density）の二つの部分を別々に積分すれば求まる。費用については、連続変数 T を離散変数に置き換えるのがコツである。硬貨の枚数は完了した半分単位の分（30 秒）が何回経過したかだけで決まるので、T がそれぞれの半分単位の区間に入る確率が求まればよい。あとは等比級数（geometric series）に類する級数が片づけてくれる。

解答

k と平均を求める

この確率密度は、$[0,1)$ 上の高さ $\frac{1}{2}$ の平らな部分と、それに続く指数関数的な裾（exponential tail）からなり、$t=1$ で跳びがある。



最初の部分が確率 $\frac{1}{2}$ を担うので、裾の部分が残りの $\frac{1}{2}$ を担わなければならない：

$$k\int_1^\infty \mathrm{e}^{-2t}\,\mathrm{d}t = k\left[-\tfrac{1}{2}\mathrm{e}^{-2t}\right]_1^\infty = \tfrac{1}{2}k\mathrm{e}^{-2} = \tfrac{1}{2}, \qquad \text{よって} \qquad k=\mathrm{e}^2.$$

（したがって $t \geq 1$ における確率密度は e^{2-2t} であり、図のとおり $t = 1$ で 1 に等しい。）

平均については、$u = t$、$\mathrm{d}v = \mathrm{e}^{-2t}\,\mathrm{d}t$ として部分積分（integration by parts）を行うと

$$\int_1^\infty t\,\mathrm{e}^{-2t}\,\mathrm{d}t = \left[-\tfrac{1}{2}t\,\mathrm{e}^{-2t}\right]_1^\infty + \tfrac{1}{2}\int_1^\infty \mathrm{e}^{-2t}\,\mathrm{d}t = \tfrac{1}{2}\mathrm{e}^{-2} + \tfrac{1}{4}\mathrm{e}^{-2} = \tfrac{3}{4}\mathrm{e}^{-2},$$

となる。ここで $t \to \infty$ のとき $t\,\mathrm{e}^{-2t} \to 0$ であることを用いた。よって

$$\mathrm{E}(T) = \int_0^1 \tfrac{1}{2}t\,\mathrm{d}t + \mathrm{e}^2 \cdot \tfrac{3}{4}\mathrm{e}^{-2} = \tfrac{1}{4} + \tfrac{3}{4} = 1,$$

であり、題意のとおり通話時間の期待値（expected value）は 1 分である。

累積分布関数

$\mathrm{F}(t) = \mathrm{P}(T \leq t)$ とおく。$t < 0$ のときは 0、$0 \leq t < 1$ のときは $\int_0^t \tfrac{1}{2}\,\mathrm{d}s = \tfrac{1}{2}t$、そして $t \geq 1$ のときは

$$\mathrm{F}(t) = \tfrac{1}{2} + \int_1^t \mathrm{e}^{2-2s}\,\mathrm{d}s = \tfrac{1}{2} + \tfrac{1}{2}(1 - \mathrm{e}^{2-2t}) = 1 - \tfrac{1}{2}\mathrm{e}^{2-2t}.$$

である。よって累積分布関数（cumulative distribution function）は

$$\mathrm{F}(t) = \begin{cases} 0 & t < 0, \\ \tfrac{1}{2}t & 0 \leq t < 1, \\ 1 - \tfrac{1}{2}\mathrm{e}^{2-2t} & t \geq 1. \end{cases}$$

となる。f は $t = 1$ で跳ぶが、F はそこで連続であり（どちらの式も $\tfrac{1}{2}$ を与える）、$t \to \infty$ のとき $\mathrm{F}(t) \to 1$ であることに注意しよう。後で使う形として、$t \geq 1$ のとき $\mathrm{P}(T > t) = \tfrac{1}{2}\mathrm{e}^{2-2t}$ である。

費用の期待値

最初に硬貨を 1 枚入れ、その後は 30 秒が経過するたびにもう 1 枚入れる。したがって通話が T 分続くとき、完了した 30 秒の回数は $\lfloor 2T \rfloor$ であり、

$$C = 10(1 + \lfloor 2T \rfloor).$$

となる。（通話がちょうど 30 秒の区切りで終わったその瞬間に硬貨を入れるかどうかは問題にならない。T は連続なので、そうなる確率は 0 である。）

考えてみよう 連続な T から C の分布へはどうやって移ればよいだろうか？

事象 $C = 10(n+1)$ はちょうど事象 $\tfrac{n}{2} \leq T < \tfrac{n+1}{2}$ であり、その確率は F から読み取れる：

- $n=0$：$\mathrm{P}(C=10)=\mathrm{F}(\frac{1}{2})=\frac{1}{4}$；
- $n=1$：$\mathrm{P}(C=20)=\mathrm{F}(1)-\mathrm{F}(\frac{1}{2})=\frac{1}{4}$；
- $n\geq 2$：両端点がともに 1 以上なので

$$\mathrm{P}\big(C=10(n+1)\big)=\tfrac{1}{2}\mathrm{e}^{2-n}-\tfrac{1}{2}\mathrm{e}^{2-(n+1)}=\tfrac{1}{2}\big(1-\mathrm{e}^{-1}\big)\mathrm{e}^{2-n}.$$

$r=\mathrm{e}^{-1}$、$m=n-2$ とおくと、$\mathrm{e}^{2-n}=r^m$、$10(n+1)=10(m+3)$ である。すると

$$\begin{aligned}\mathrm{E}(C)&=10\cdot\tfrac{1}{4}+20\cdot\tfrac{1}{4}+\sum_{m=0}^{\infty}10(m+3)\cdot\tfrac{1}{2}(1-r)\,r^m\\&=\tfrac{15}{2}+5(1-r)\left(\sum_{m=0}^{\infty}mr^m+3\sum_{m=0}^{\infty}r^m\right).\end{aligned}$$

$0<r<1$ なので、等比級数から $\sum r^m=\dfrac{1}{1-r}$ が得られ、これを項別微分（differentiating term by term）して r を掛ければ $\sum mr^m=\dfrac{r}{(1-r)^2}$ が得られる。したがって

$$\mathrm{E}(C)=\tfrac{15}{2}+5(1-r)\left(\frac{r}{(1-r)^2}+\frac{3}{1-r}\right)=\tfrac{15}{2}+\frac{5r}{1-r}+15.$$

最後に $\dfrac{r}{1-r}=\dfrac{\mathrm{e}^{-1}}{1-\mathrm{e}^{-1}}=\dfrac{1}{\mathrm{e}-1}$ なので

$$\mathrm{E}(C)=22\tfrac{1}{2}+\frac{5}{\mathrm{e}-1},$$

となり、題意が示された。数値ではおよそ 25.41 ペンスである。

考察

1 分あたり 20 ペンスならば、平均 1 分の通話の費用は平均 20 ペンスになるはずだが、実際には $\mathrm{E}(C)\approx 25.4$ である。この差が生じるのは、その男性が 30 秒ごとに**前払い**しているからである。どの硬貨も 30 秒分をまるごと買うが、最後の 1 枚分が使い切られることはほとんどない。実際、$\lfloor 2T\rfloor>2T-1$ なので、すべての通話について $C=10(1+\lfloor 2T\rfloor)>20T$ が成り立つ。費用は通話時間を次の 30 秒の区切りまで**切り上げる**ので、$\mathrm{E}(C)$ は $20\,\mathrm{E}(T)$ を、支払ったのに使わなかった時間の期待値の 20 倍（ここではおよそ 5.4 ペンス）だけ上回る。

検算

- C の確率の和は 1 になる：$\frac{1}{4}+\frac{1}{4}+\sum_{m\geq 0}\frac{1}{2}(1-r)r^m=\frac{1}{2}+\frac{1}{2}=1$。

- 評価 $20T < C \leq 20T + 10$ から $20 < \mathrm{E}(C) \leq 30$ が得られ、25.41 は余裕をもってその範囲に収まる。最後の 30 秒のうち使われない時間が一様分布（uniform）に従うなら、超過分はおよそ 5 ペンスと予想されるが、長い裾のためにそれより少し多くなる。
- 数値的には、シンプソンの公式（Simpson's rule）により全確率 1、平均 1 が得られ、また 10^6 回の通話のシミュレーション（半分は $[0,1)$ 上の一様分布、残り半分は 1 にレート 2 の指数分布を加えたもの）では平均費用 25.41 ペンスとなり、$22.5 + 5/(\mathrm{e}-1) = 25.4099$ と一致した。

よくある間違い

- 最初の硬貨を忘れて、硬貨の枚数を $1 + \lfloor 2T \rfloor$ ではなく $\lfloor 2T \rfloor$ としてしまう。これでは答えがちょうど 10 ペンス少なくなる。
- $t = \frac{1}{2}$ で F に $t \geq 1$ の式を使ってしまう。区間 $[0, \frac{1}{2})$ と $[\frac{1}{2}, 1)$ は一様な部分を使って扱わなければならない。

STEP のその先へ

T の裾は無記憶性（memoryless property）をもつことに注目しよう。$T \geq 1$ という条件の下で、残りの時間 $T - 1$ はレート 2 の指数分布に従う。$n \geq 2$ に対する確率が等比数列をなすのはこのためである。$N = \lfloor 2T \rfloor$ とおくと、裾和公式（tail-sum formula）$\mathrm{E}(N) = \sum_{n \geq 1} \mathrm{P}(N \geq n)$ を使えば同じ答えがさらに速く得られる。このような離散的な振る舞いと連続的な振る舞いの混合は、Grimmett and Welsh, *Probability: An Introduction*（確率論入門）で丁寧に扱われており、Blitzstein and Hwang, *Introduction to Probability*（確率入門）には無記憶性の例がさらに数多く載っている。

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。