詳細解答

STEP I 樣本試卷，第 16 題

概率與統計 • 連續隨機變量

> **STEP I 樣本試卷，第 16 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/spec-s1-q16

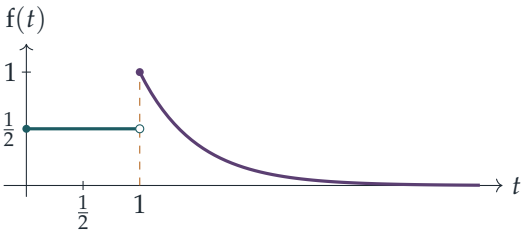
關鍵思路

未知數 k 由「總概率為 1」這一事實決定，其餘各項只要把密度函數的兩段分開積分便可得出。至於費用，竅門是把連續變量 T 化為離散變量：硬幣的數目只取決於已過了多少個完整的半分鐘，所以我們需要求 T 落在每個半分鐘區間內的概率。餘下的工作交給一個幾何型的級數便可完成。

解答

求 k 及平均值

密度函數在 $[0,1)$ 上是高度為 $\frac{1}{2}$ 的水平線段，其後接上一條指數尾部（exponential tail），並在 $t=1$ 處有一個跳躍。



第一段貢獻的概率為 $\frac{1}{2}$，所以尾部必須貢獻餘下的 $\frac{1}{2}$：

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

（因此在 $t \geq 1$ 上的密度函數為 e^{2-2t}，它在 $t=1$ 處等於 1，與草圖相符。）

至於平均值，取 $u = t$、$\mathrm{d}v = \mathrm{e}^{-2t}\,\mathrm{d}t$ 作分部積分（integration by parts）：

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

其中用到當 $t \to \infty$ 時 $t\,\mathrm{e}^{-2t} \to 0$。因此

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

所以通話時間的期望值（expected value）是一分鐘，正如題目所求。

累積分佈函數

記 $\mathrm{F}(t) = \mathrm{P}(T \le t)$，即累積分佈函數（cumulative distribution function）。當 $t < 0$ 時它等於 0；當 $0 \le t < 1$ 時它等於 $\int_0^t \frac{1}{2}\,\mathrm{d}s = \frac{1}{2}t$；而當 $t \ge 1$ 時，

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

所以

$$\mathrm{F}(t) = \begin{cases} 0 & t < 0, \\ \frac{1}{2}t & 0 \le t < 1, \\ 1 - \frac{1}{2}\mathrm{e}^{2-2t} & t \ge 1. \end{cases}$$

留意雖然 f 在 $t = 1$ 處有跳躍，F 在該處卻是連續的（兩條公式都給出 $\frac{1}{2}$），而且當 $t \to \infty$ 時 $\mathrm{F}(t) \to 1$。下文會用到這個形式：當 $t \ge 1$ 時，$\mathrm{P}(T > t) = \frac{1}{2}\mathrm{e}^{2-2t}$。

費用的期望值

通話開始時投入一枚硬幣，其後每過完一個半分鐘再投入一枚。所以若通話歷時 T 分鐘，完整的半分鐘數目為 $\lfloor 2T \rfloor$，而

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

（若通話恰好在某個半分鐘的整點結束，那一刻有沒有投入硬幣並不要緊：T 是連續的，所以這種情況的概率為 0。）

想一想 怎樣由連續的 T 得出 C 的分佈？

事件 $C = 10(n+1)$ 正正就是事件 $\frac{n}{2} \le T < \frac{n+1}{2}$，其概率可直接由 F 讀出：

- $n = 0$：$\mathrm{P}(C = 10) = \mathrm{F}(\frac{1}{2}) = \frac{1}{4}$；
- $n = 1$：$\mathrm{P}(C = 20) = \mathrm{F}(1) - \mathrm{F}(\frac{1}{2}) = \frac{1}{4}$；

- $n \geq 2$：兩個端點都不小於 1，所以

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

設 $r = \mathrm{e}^{-1}$ 及 $m = n-2$，則 $\mathrm{e}^{2-n} = r^m$ 而 $10(n+1) = 10(m+3)$。於是

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

由於 $0 < r < 1$，由幾何級數（geometric series）得 $\sum r^m = \dfrac{1}{1-r}$，把它逐項求導（再乘以 r）便得 $\sum m r^m = \dfrac{r}{(1-r)^2}$。因此

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

最後，$\dfrac{r}{1-r} = \dfrac{\mathrm{e}^{-1}}{1-\mathrm{e}^{-1}} = \dfrac{1}{\mathrm{e}-1}$，所以

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

正如題目所求。其數值約為 25.41 便士。

評論

若按每分鐘 20 便士計算，平均長度為一分鐘的通話平均應花費 20 便士，但 $\mathrm{E}(C) \approx 25.4$。出現這個差距，是因為那人要為每個半分鐘預先付款：每枚硬幣買下完整的半分鐘，但最後一枚幾乎從來不會用盡。事實上，由於 $\lfloor 2T \rfloor > 2T - 1$，每一次通話都有 $C = 10(1 + \lfloor 2T \rfloor) > 20T$；費用把通話長度向上捨入至下一個半分鐘，所以 $\mathrm{E}(C)$ 比 $20\,\mathrm{E}(T)$ 多出的數額，等於已付款而未用的時間的期望值的二十倍（在這裏約為 5.4 便士）。

驗算

- C 的各個概率之和為 1：$\frac{1}{4} + \frac{1}{4} + \sum_{m\geq 0} \frac{1}{2}(1-r)r^m = \frac{1}{2} + \frac{1}{2} = 1$。
- 由界限 $20T < C \leq 20T + 10$ 得 $20 < \mathrm{E}(C) \leq 30$，而 25.41 穩妥地落在其中。假如最後一個半分鐘內未用的時間是均勻分佈（uniform distribution）的，我們會預期多付約 5 便士；長尾令這個數額稍為多一點。
- 數值上，辛普森法則（Simpson's rule）給出總概率為 1、平均值為 1；另外模擬 10^6 次通話（一半在 $[0,1)$ 上均勻分佈，一半為 1 加上率為 2 的指數分佈），所得平均費用為 25.41 便士，與 $22.5 + 5/(\mathrm{e}-1) = 25.4099$ 吻合。

常見錯誤

- 把硬幣數目當作 $\lfloor 2T \rfloor$ 而非 $1 + \lfloor 2T \rfloor$，忘記了開始時投入的那一枚；這樣答案會剛好少了 10 便士。
- 在 $t = \frac{1}{2}$ 處套用 F 在 $t \geq 1$ 時的公式：區間 $[0, \frac{1}{2})$ 和 $[\frac{1}{2}, 1)$ 必須用均勻分佈的那一段來處理。

STEP 之外

留意 T 的尾部具有無記憶性（memoryless property）：已知 $T \geq 1$ 時，剩餘時間 $T - 1$ 服從率為 2 的指數分佈，這正是 $n \geq 2$ 時各個概率構成幾何數列（geometric sequence）的原因。記 $N = \lfloor 2T \rfloor$，則尾和公式（tail-sum formula）$\mathrm{E}(N) = \sum_{n \geq 1} \mathrm{P}(N \geq n)$ 能更快地得出相同的答案。像這樣混合了離散與連續特性的分佈，在 Grimmett 與 Welsh 的 *Probability: An Introduction* 中有仔細的論述，而 Blitzstein 與 Hwang 的 *Introduction to Probability* 則載有更多關於無記憶性的例子。

Beyond STEP 詳細解答：https://beyondstep.keithmcnulty.org/zh-hant/。解答文字採用 CC0 1.0；題目本身版權屬 Cambridge University Press & Assessment 所有（©），本文件不予轉載。由機器從英文翻譯而成；尚未經審閱。