详细解答

STEP I 样本试卷，第 16 题

概率与统计 · 连续随机变量

STEP I 样本试卷，第 16 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/spec-s1-q16

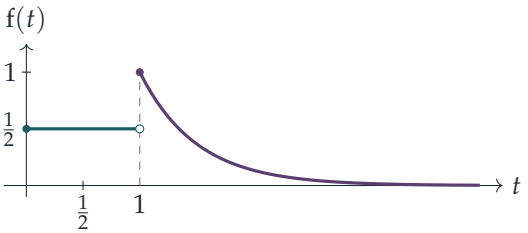
关键思路

未知数 k 由“总概率为 1”这一事实决定，其余各项只要把密度函数的两段分开积分便可得出。至于费用，窍门是把连续变量 T 化为离散变量：硬币的数目只取决于已过了多少个完整的半分钟，所以我们需要求 T 落在每个半分钟区间内的概率。余下的工作交给一个几何型的级数便可完成。

解答

求 k 及平均值

密度函数在 $[0,1)$ 上是高度为 $\frac{1}{2}$ 的水平线段，其后接上一条指数尾部（exponential tail），并在 $t=1$ 处有一个跳跃。



第一段贡献的概率为 $\frac{1}{2}$，所以尾部必须贡献余下的 $\frac{1}{2}$：

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

（因此在 $t \geq 1$ 上的密度函数为 e^{2-2t}，它在 $t=1$ 处等于 1，与草图相符。）

至于平均值，取 $u = t$、$\mathrm{d}v = \mathrm{e}^{-2t}\,\mathrm{d}t$ 作分部积分（integration by parts）：

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

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

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

所以通话时间的期望值（expected value）是一分钟，正如题目所求。

累积分布函数

记 $\mathrm{F}(t) = \mathrm{P}(T \le t)$，即累积分布函数（cumulative distribution function）。当 $t < 0$ 时它等于 0；当 $0 \le t < 1$ 时它等于 $\int_0^t \tfrac{1}{2}\,\mathrm{d}s = \tfrac{1}{2}t$；而当 $t \ge 1$ 时，

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

所以

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

留意虽然 f 在 $t = 1$ 处有跳跃，F 在该处却是连续的（两条公式都给出 $\tfrac{1}{2}$），而且当 $t \to \infty$ 时 $\mathrm{F}(t) \to 1$。下文会用到这个形式：当 $t \ge 1$ 时，$\mathrm{P}(T > t) = \tfrac{1}{2}\mathrm{e}^{2-2t}$。

费用的期望值

通话开始时投入一枚硬币，其后每过完一个半分钟再投入一枚。所以若通话历时 T 分钟，完整的半分钟数目为 $\lfloor 2T \rfloor$，而

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

（若通话恰好在某个半分钟的整点结束，那一刻有没有投入硬币并不要紧：T 是连续的，所以这种情况的概率为 0。）

> **想一想** 怎样由连续的 T 得出 C 的分布？

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

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

- $n \geq 2$：两个端点都不小于 1，所以

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

设 $r = \mathrm{e}^{-1}$ 及 $m = n-2$，则 $\mathrm{e}^{2-n} = r^m$ 而 $10(n+1) = 10(m+3)$。于是

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

由于 $0 < r < 1$，由几何级数（geometric series）得 $\sum r^m = \dfrac{1}{1-r}$，把它逐项求导（再乘以 r）便得 $\sum m r^m = \dfrac{r}{(1-r)^2}$。因此

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

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

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

正如题目所求。其数值约为 25.41 便士。

评论

若按每分钟 20 便士计算，平均长度为一分钟的通话平均应花费 20 便士，但 $\mathrm{E}(C) \approx 25.4$。出现这个差距，是因为那人要为每个半分钟预先付款：每枚硬币买下完整的半分钟，但最后一枚几乎从来不会用尽。事实上，由于 $\lfloor 2T \rfloor > 2T - 1$，每一次通话都有 $C = 10(1 + \lfloor 2T \rfloor) > 20T$；费用把通话长度向上舍入至下一个半分钟，所以 $\mathrm{E}(C)$ 比 $20\,\mathrm{E}(T)$ 多出的数额，等于已付款而未用的时间的期望值的二十倍（在这里约为 5.4 便士）。

验算

- C 的各个概率之和为 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 稳妥地落在其中。假如最后一个半分钟内未用的时间是均匀分布（uniform distribution）的，我们会预期多付约 5 便士；长尾令这个数额稍为多一点。
- 数值上，辛普森法则（Simpson's rule）给出总概率为 1、平均值为 1；另外模拟 10^6 次通话（一半在 $[0,1)$ 上均匀分布，一半为 1 加上率为 2 的指数分布），所得平均费用为 25.41 便士，与 $22.5 + 5/(\mathrm{e}-1) = 25.4099$ 吻合。

常见错误

- 把硬币数目当作 $\lfloor 2T \rfloor$ 而非 $1 + \lfloor 2T \rfloor$，忘记了开始时投入的那一枚；这样答案会刚好少了 10 便士。
- 在 $t = \frac{1}{2}$ 处套用 F 在 $t \geq 1$ 时的公式：区间 $[0, \frac{1}{2})$ 和 $[\frac{1}{2}, 1)$ 必须用均匀分布的那一段来处理。

STEP 之外

留意 T 的尾部具有无记忆性（memoryless property）：已知 $T \geq 1$ 时，剩余时间 $T - 1$ 服从率为 2 的指数分布，这正是 $n \geq 2$ 时各个概率构成几何数列（geometric sequence）的原因。记 $N = \lfloor 2T \rfloor$，则尾和公式（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。像这样混合了离散与连续特性的分布，在 Grimmett 与 Welsh 的 *Probability: An Introduction* 中有仔细的论述，而 Blitzstein 与 Hwang 的 *Introduction to Probability* 则载有更多关于无记忆性的例子。

Beyond STEP 详细解答：https://beyondstep.keithmcnulty.org/zh-hans/。解答文字采用 CC0 1.0；题目本身版权属 Cambridge University Press & Assessment 所有（©），本文件不予转载。由机器从英文翻译而成；尚未经审阅。