詳細解答

STEP III 2003，第 12 題

概率與統計 • 指數分佈的等候時間

STEP III 2003，第 12 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/03-s3-q12

關鍵思路

如果任何一段時間內的打擾次數都服從泊松分佈（Poisson distribution），那麼「下一次打擾在 t 分鐘之後才出現」與「在接下來的 t 分鐘內完全沒有打擾」是同一個事件，這樣就把離散分佈（discrete distribution）變成連續分佈（continuous distribution）：等候時間服從指數分佈（exponential distribution）。其後，相繼兩次打擾之間的各個間隔互相獨立（independent），所以尋找一個長間隔的過程，就是一連串獨立試驗，每次試驗不是「成功」（間隔達半小時或以上）就是「失敗」。餘下的工作交給幾何分佈（geometric distribution）和條件期望（conditional expectation）便可。

解答

T 的分佈

打擾平均每十分鐘出現一次，所以發生率是每分鐘 $\frac{1}{10}$ 次，而在長 t 分鐘的時段內，打擾次數服從平均值為 $\frac{t}{10}$ 的泊松分佈。

從剛剛發生的那次打擾開始計時。到下一次打擾的時間 T 不少於 t，當且僅當在接下來的 t 分鐘內沒有打擾，所以

$$\mathrm{P}(T \geq t) = \mathrm{P}(\text{沒有任何打擾，為時} t) = \mathrm{e}^{-t/10}.$$

因此，對於 $t \geq 0$，工作時間少於 t 分鐘的概率是

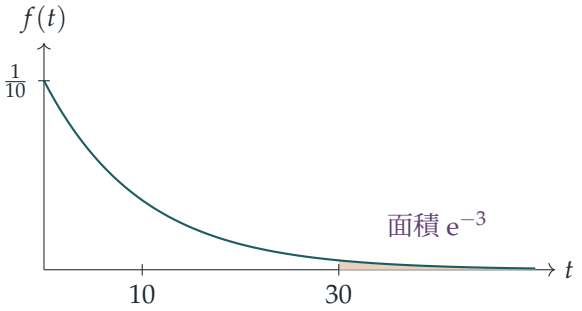
$$\mathrm{P}(T < t) = 1 - \mathrm{e}^{-t/10}.$$

這就是 T 的累積分佈函數（cumulative distribution function）$F(t)$（當 $t < 0$ 時 $F(t) = 0$）。

求導，得

$$f(t) = F'(t) = \tfrac{1}{10}\,\mathrm{e}^{-t/10} \quad (t \geq 0), \qquad f(t) = 0 \quad (t < 0).$$

所以 T 服從平均值為 10 分鐘的指數分佈。



打擾次數的期望值

由於這個過程沒有記憶（互不重疊的時段內的泊松計數互相獨立），每當一次打擾發生，到下一次打擾的時間都重新開始計算，其分佈與 T 相同，並且獨立於之前發生的一切。所以相繼的間隔 $T_1, T_2, \ldots$ 互相獨立，各自的分佈都與 T 相同。

如果一個間隔不少於 30 分鐘，就稱它為長間隔。每個間隔是長間隔的概率為

$$p = \mathrm{P}(T \geq 30) = \mathrm{e}^{-30/10} = \mathrm{e}^{-3},$$

而且各個間隔互相獨立。設 N 為第一個長間隔出現之前的打擾次數（剛剛發生的那一次不計算在內）。每個短間隔都以一次打擾告終，而第一個長間隔就是我那不受打擾的半小時，所以 N 就是第一個長間隔之前的短間隔數目。因此

$$\mathrm{P}(N = n) = (1-p)^n p, \qquad n = 0, 1, 2, \ldots$$

? **想一想** 這個幾何分佈由 0 而不是由 1 開始，我們怎樣求它的平均值？

對幾何級數（geometric series）$\sum_{n\geq 0} x^n = \frac{1}{1-x}$ 求導，可得當 $|x| < 1$ 時 $\sum_{n\geq 1} n x^{n-1} = \frac{1}{(1-x)^2}$。取 $x = 1-p$，

$$\mathrm{E}(N) = \sum_{n=1}^{\infty} n(1-p)^n p = p(1-p)\cdot\frac{1}{p^2} = \frac{1-p}{p}.$$

（另一個做法：$N+1$ 是直至首次成功為止（包括該次）的試驗次數，其平均值為 $\frac{1}{p}$。）代入 $p = \mathrm{e}^{-3}$，

$$\mathrm{E}(N) = \frac{1-\mathrm{e}^{-3}}{\mathrm{e}^{-3}} = \mathrm{e}^3 - 1,$$

正是所要求的結果。

短間隔長度的期望值

我們要求 $\mathrm{E}(T \mid T < 30)$。已知 $T < 30$ 時，T 的條件概率密度函數（conditional density）在 $0 \le t < 30$ 上是 $f(t)/\mathrm{P}(T < 30)$，所以

$$\mathrm{E}(T \mid T < 30) = \frac{1}{1 - \mathrm{e}^{-3}} \int_0^{30} \frac{t}{10}\, \mathrm{e}^{-t/10}\, \mathrm{d}t.$$

用分部積分法（integration by parts），取 $u = t$ 和 $\mathrm{d}v = \frac{1}{10}\mathrm{e}^{-t/10}\,\mathrm{d}t$，於是 $v = -\mathrm{e}^{-t/10}$，

$$\begin{aligned}\int_0^{30} \frac{t}{10}\, \mathrm{e}^{-t/10}\, \mathrm{d}t &= \Big[-t\, \mathrm{e}^{-t/10} \Big]_0^{30} + \int_0^{30} \mathrm{e}^{-t/10}\, \mathrm{d}t \\ &= -30\, \mathrm{e}^{-3} + 10(1 - \mathrm{e}^{-3}) \\ &= 10 - 40\, \mathrm{e}^{-3}.\end{aligned}$$

因此，把分子和分母同乘以 e^3，

$$\mathrm{E}(T \mid T < 30) = \frac{10 - 40\, \mathrm{e}^{-3}}{1 - \mathrm{e}^{-3}} = \frac{10\, \mathrm{e}^3 - 40}{\mathrm{e}^3 - 1} \text{ 分鐘} \approx 8.43 \text{ 分鐘}.$$

這個數值略小於 10，是合理的：我們已經把長間隔剔除了。

等候時間的期望值

在不受打擾的半小時開始之前，等候時間就是 N 個短間隔之和。已知 $N = n$ 時，這是 n 個獨立的間隔，每個都以短於 30 分鐘為條件，所以已知 $N = n$ 時等候時間的期望值是 $n\,\mathrm{E}(T \mid T < 30)$。再對 N 取平均，

$$\mathrm{E}(\text{等候時間}) = \mathrm{E}(N)\, \mathrm{E}(T \mid T < 30) = (\mathrm{e}^3 - 1) \cdot \frac{10\, \mathrm{e}^3 - 40}{\mathrm{e}^3 - 1} = 10\, \mathrm{e}^3 - 40 \text{ 分鐘},$$

即大約 161 分鐘，或 2 小時 41 分鐘。（如果想求直至試卷完成為止的時間，就要加上那半小時本身：$10\, \mathrm{e}^3 - 10 \approx 191$ 分鐘。）

驗算

- $\int_0^{\infty} \frac{1}{10}\mathrm{e}^{-t/10}\,\mathrm{d}t = 1$，所以 f 確實是一個概率密度函數，而它的平均值是 10，與「平均每十分鐘一次」吻合。
- 在沒有條件的情況下，$\mathrm{E}(T) = 10$ 應該可以分拆為 $\mathrm{P}(T < 30)\, \mathrm{E}(T \mid T < 30) + \mathrm{P}(T \ge 30)\, \mathrm{E}(T \mid T \ge 30)$。由無記憶性（memorylessness）可知 $\mathrm{E}(T \mid T \ge 30) = 40$，而果然有 $(10 - 40\mathrm{e}^{-3}) + 40\mathrm{e}^{-3} = 10$。
- 把這個過程模擬運行 200 000 次，得出平均打擾次數為 19.1（對照 $\mathrm{e}^3 - 1 \approx 19.09$），平均等候時間為 161.1 分鐘（對照 $10\mathrm{e}^3 - 40 \approx 160.86$）。

常見錯誤

- 使用 $\mathrm{E}(N) = \frac{1}{p} = \mathrm{e}^3$：這樣連長間隔也計算在內，但長間隔結束時，那半小時之前並沒有再出現打擾。
- 把 $\mathrm{E}(N)$ 乘以無條件平均值 10 來求等候時間。組成等候時間的間隔已知是短間隔，所以它們的平均值是 $\mathrm{E}(T \mid T < 30)$，而不是 10。

STEP 之外

這些打擾構成一個泊松過程（*Poisson process*），而各個間隔是互相獨立的指數分佈變量這個結果，正是泊松過程的基本性質之一；最後一步 $\mathrm{E}(\text{共}N\text{ 項之和}) = \mathrm{E}(N) \times \mathrm{E}(\text{每項})$ 是一個取條件的論證，與沃爾德恆等式（*Wald's identity*）關係密切。兩者合起來，通向更新理論（renewal theory）和排隊模型（queueing models），這些都是大學一、二年級概率課程的內容。Grimmett 與 Welsh 合著的 *Probability: An Introduction*（概率論入門）以合適的程度講解泊松過程和條件期望。

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。