詳細解答

STEP I 2016，第 13 題

概率與統計 • 指數分佈的順序統計量

STEP I 2016，第 13 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/16-s1-q13

關鍵思路

對於第一封和第二封郵件的到達，要問的是它們何時尚未發生。第一封電子郵件在時間 t 之後才到達，當且僅當全部 n 封都在 t 之後到達；第二封在 t 之後才到達，當且僅當到那時至多只有一封已到達。利用獨立性（independence），這兩個事件都很容易計算，再把生存概率（survival probability）求導，便得到概率密度函數（density）。

解答

設 $X_1, \ldots, X_n$ 為各封郵件的到達時間。它們互相獨立，而且每一個都滿足

$$\mathrm{P}(X_i > t) = \int_t^\infty \lambda \mathrm{e}^{-\lambda s}\,\mathrm{d}s = \mathrm{e}^{-\lambda t}, \qquad \mathrm{P}(X_i \le t) = 1 - \mathrm{e}^{-\lambda t}$$

其中 $t \ge 0$。我們會用到標準積分 $\int_0^\infty t k\mathrm{e}^{-kt}\,\mathrm{d}t = \frac{1}{k}$（$k > 0$），這正是參數為 k 的指數分佈（exponential distribution）的平均值。（用分部積分法（integration by parts）：$[-t\mathrm{e}^{-kt}]_0^\infty + \int_0^\infty \mathrm{e}^{-kt}\,\mathrm{d}t = 0 + \frac{1}{k}$。）

第 (i) 部分

第一封郵件在 t 之後才到達，當且僅當每一封郵件都在 t 之後才到達。由獨立性，

$$\mathrm{P}(T > t) = \prod_{i=1}^{n} \mathrm{P}(X_i > t) = \left(\mathrm{e}^{-\lambda t}\right)^n = \mathrm{e}^{-n\lambda t}.$$

所以累積分佈函數（cumulative distribution function）為 $F(t) = 1 - \mathrm{e}^{-n\lambda t}$，其中 $t \ge 0$。求導便得到密度函數

$$f(t) = F'(t) = n\lambda\,\mathrm{e}^{-n\lambda t} \qquad (t \ge 0),$$

正如所求。這是參數為 $n\lambda$ 的指數密度函數，所以取 $k = n\lambda$ 代入標準積分，得

$$\mathrm{E}(T) = \frac{1}{n\lambda}.$$

第 (ii) 部分

設 S 為第二封郵件的到達時間（因此需要 $n \geq 2$）。第二封郵件在 t 之後才到達，當且僅當到時間 t 為止至多一封郵件已到達：要麼一封也沒有，要麼恰好一封。

? **想一想** 到時間 t 為止恰好有一封郵件已到達，有多少種可能情況？

n 封郵件中任何一封都可以是那封提早到達的。那一封已到達（概率為 $1 - \mathrm{e}^{-\lambda t}$），而其餘 $n-1$ 封尚未到達（每封概率為 $\mathrm{e}^{-\lambda t}$）。由獨立性，並且由於這 n 種情況互斥（mutually exclusive），

$$\mathrm{P}(S > t) = \mathrm{e}^{-n\lambda t} + n\left(1 - \mathrm{e}^{-\lambda t}\right)\mathrm{e}^{-(n-1)\lambda t}.$$

展開第二項得 $n\mathrm{e}^{-(n-1)\lambda t} - n\mathrm{e}^{-n\lambda t}$，所以

$$\mathrm{P}(S > t) = n\,\mathrm{e}^{-(n-1)\lambda t} - (n-1)\,\mathrm{e}^{-n\lambda t}.$$

簡單驗證一下：當 $t = 0$ 時它等於 1，而當 $t \to \infty$ 時它趨於 0。

S 的密度函數 g 是 $1 - \mathrm{P}(S > t)$ 的導數：

$$\begin{aligned} g(t) &= -\frac{\mathrm{d}}{\mathrm{d}t}\left(n\,\mathrm{e}^{-(n-1)\lambda t} - (n-1)\,\mathrm{e}^{-n\lambda t}\right) \\ &= n(n-1)\lambda\,\mathrm{e}^{-(n-1)\lambda t} - n(n-1)\lambda\,\mathrm{e}^{-n\lambda t} \\ &= n(n-1)\lambda\left(\mathrm{e}^{-(n-1)\lambda t} - \mathrm{e}^{-n\lambda t}\right) \qquad (t \geq 0). \end{aligned}$$

為求期望值（expectation），把 g 寫成指數密度函數的組合：

$$g(t) = n \cdot (n-1)\lambda\,\mathrm{e}^{-(n-1)\lambda t} \;-\; (n-1)\cdot n\lambda\,\mathrm{e}^{-n\lambda t}.$$

分別取 $k = (n-1)\lambda$ 和 $k = n\lambda$ 代入標準積分（這是成立的，因為 $n \geq 2$，所以兩種情況下都有 $k > 0$），

$$\begin{aligned} \mathrm{E}(S) = \int_0^\infty t\,g(t)\,\mathrm{d}t &= \frac{n}{(n-1)\lambda} - \frac{n-1}{n\lambda} \\ &= \frac{1}{\lambda}\cdot\frac{n^2 - (n-1)^2}{n(n-1)} = \frac{1}{\lambda}\cdot\frac{2n-1}{n(n-1)}. \end{aligned}$$

由於 $\dfrac{1}{n-1} + \dfrac{1}{n} = \dfrac{n + (n-1)}{n(n-1)} = \dfrac{2n-1}{n(n-1)}$，所以

$$\mathrm{E}(S) = \frac{1}{\lambda}\left(\frac{1}{n-1} + \frac{1}{n}\right),$$

這正是所求的結果。

想一想 為甚麼答案是兩個簡單項之和？

指數分佈具有無記憶性（memoryless）。第一封郵件在一段速率為 $n\lambda$ 的指數時間後到達，平均值為 $\frac{1}{n\lambda}$。之後，其餘 $n-1$ 封郵件重新開始，就好像時鐘剛被重設一樣，所以額外的等待時間是速率為 $(n-1)\lambda$ 的指數分佈，平均值為 $\frac{1}{(n-1)\lambda}$。把兩段等待時間相加，立即得到答案，這也是記住它的好方法。

驗算

- 當 $n=2$ 且 $\lambda=1$ 時，第二封郵件的到達時間是兩個獨立指數時間中較晚的一個。它的平均值應為 $1+\frac{1}{2}=1.5$，而以電腦模擬 10^5 次試驗得到 1.501。
- 密度函數 g 的積分為 1：$\int_0^\infty g = n-(n-1)=1$。
- 當 $n=5$ 且 $\lambda=1$ 時，公式給出 $\mathrm{E}(S)=\frac{1}{4}+\frac{1}{5}=0.45$，與 $\int_0^\infty t\,g(t)\,\mathrm{d}t$ 的數值計算結果吻合。

常見錯誤

- 在恰好一封郵件已到達的情況中漏掉因子 n。n 封中任何一封都可以是提早到達的那封。
- 把 $\mathrm{P}(S>t)$ 寫成恰好一封已到達的概率，這樣便漏掉了一封也未到達的情況。

STEP 之外

T 和 S 是一個指數樣本的首兩個順序統計量（order statistics）。上面的無記憶性論證可以推廣，證明第 k 封郵件的到達時間平均值為 $\frac{1}{\lambda}\left(\frac{1}{n}+\frac{1}{n-1}+\cdots+\frac{1}{n-k+1}\right)$，而這正是泊松過程（Poisson process）和排隊論（queueing theory）的基礎。Blitzstein 和 Hwang 的 *Introduction to Probability*（《概率導論》）對順序統計量和無記憶性都有清晰的講解。

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。