解答例

STEP I 2016、第 13 問

確率・統計 ・ 指数分布の順序統計量

> **STEP I 2016、問 13**
>
> 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/16-s1-q13

鍵となる考え方

1 通目と 2 通目の到着については、それらがまだ起きていないのはいつかを考える。1 通目のメールが時刻 t より後に届くのは、n 通すべてがそうなるとき、かつそのときに限る。また 2 通目が t より後に届くのは、その時点までに届いたメールが高々 1 通であるとき、かつそのときに限る。どちらの事象も独立性（independence）を使えば容易に数えられ、生存確率（survival probability）を微分すれば確率密度関数（density）が得られる。

解答

$X_1, \ldots, X_n$ を到着時刻とする。これらは独立であり、それぞれについて

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

が $t \ge 0$ に対して成り立つ。$k > 0$ に対する標準的な積分 $\int_0^\infty t\, k\mathrm{e}^{-kt}\,\mathrm{d}t = \frac{1}{k}$ を用いる。これはパラメータ k の指数分布（exponential distribution）の平均（mean）である。（部分積分（integration by parts）すればよい：$[-t\mathrm{e}^{-kt}]_0^\infty + \int_0^\infty \mathrm{e}^{-kt}\,\mathrm{d}t = 0 + \frac{1}{k}$。）

小問 (i)

1 通目の到着が t より遅いのは、すべてのメールが t より遅く届くとき、かつそのときに限る。独立性より

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

よって累積分布関数（cumulative distribution function）は $t \geq 0$ に対して $F(t) = 1 - \mathrm{e}^{-n\lambda t}$ である。微分すると確率密度関数

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

が得られ、これが示すべきものであった。これはパラメータ $n\lambda$ の指数分布の密度関数であるから、$k = n\lambda$ として標準的な積分を用いると

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

小問 (ii)

S を 2 通目のメールの到着時刻とする（したがって $n \geq 2$ が必要である）。2 通目の到着が t より遅いのは、時刻 t までに届いたメールが**高々 1 通**であるとき、すなわち 1 通も届いていないか、ちょうど 1 通だけ届いているとき、かつそのときに限る。

> **考えてみよう** 時刻 t までにちょうど 1 通だけ届いている場合は何通りあるか？

n 通のメールのうち、どれが早く届いた 1 通であってもよい。その 1 通は届いており（確率 $1 - \mathrm{e}^{-\lambda t}$）、残りの $n - 1$ 通は届いていない（それぞれ確率 $\mathrm{e}^{-\lambda t}$）。独立性と、n 個の場合が互いに排反（mutually exclusive）であることから、

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

第 2 項を展開すると $n\mathrm{e}^{-(n-1)\lambda t} - n\mathrm{e}^{-n\lambda t}$ となるので、

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

簡単な妥当性の確認として、これは $t = 0$ で 1 に等しく、$t \to \infty$ のとき 0 に近づく。

S の密度関数 g は $1 - \mathrm{P}(S > t)$ の導関数である：

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

期待値（expectation）を求めるため、g を指数分布の密度関数の組合せとして書く：

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

$k = (n-1)\lambda$ と $k = n\lambda$ として標準的な積分を用いると（$n \geq 2$ なのでどちらの場合も

$k > 0$ であり、適用できる)、

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

$\dfrac{1}{n-1} + \dfrac{1}{n} = \dfrac{n + (n-1)}{n(n-1)} = \dfrac{2n-1}{n(n-1)}$ であるから、これは

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

となり、示すべき結果が得られた。

考えてみよう なぜ答えは 2 つの単純な項の和になるのか？

指数分布は無記憶性（memoryless property）をもつ。1 通目は率（rate）$n\lambda$ の指数分布に従う時間の後に届き、その平均は $\frac{1}{n\lambda}$ である。その後、残りの $n-1$ 通のメールは、まるで時計がちょうどリセットされたかのように新たに始まるので、追加の待ち時間は率 $(n-1)\lambda$、平均 $\frac{1}{(n-1)\lambda}$ の指数分布に従う。2 つの待ち時間を足せばただちに答えが得られ、これは結果を覚えるよい方法でもある。

検算

- $n = 2$ かつ $\lambda = 1$ のとき、2 通目の到着は独立な 2 つの指数分布に従う時間のうち遅いほうである。その平均は $1 + \frac{1}{2} = 1.5$ となるはずであり、10^5 回の試行によるコンピュータシミュレーションでは 1.501 が得られた。
- 密度関数 g の積分は 1 になる：$\int_0^\infty g = n - (n-1) = 1$。
- $n = 5$ かつ $\lambda = 1$ のとき、公式から $\mathrm{E}(S) = \frac{1}{4} + \frac{1}{5} = 0.45$ が得られ、これは $\int_0^\infty t\, g(t)\, \mathrm{d}t$ の数値計算と一致する。

よくある間違い

- ちょうど 1 通が届いている場合の因子 n を忘れる。n 通のどれが早く届いた 1 通であってもよい。
- $\mathrm{P}(S > t)$ を、ちょうど 1 通が届いている確率として書いてしまう。これでは 1 通も届いていない場合が抜け落ちる。

STEP のその先へ

T と S は、指数分布に従う標本の最初の 2 つの**順序統計量**（order statistics）である。上の無記憶性による議論を拡張すると、k 番目の到着の平均が $\frac{1}{\lambda}\left(\frac{1}{n}+\frac{1}{n-1}+\cdots+\frac{1}{n-k+1}\right)$ であることが示され、これはポアソン過程（Poisson process）や待ち行列理論（queueing theory）の基礎となっている。Blitzstein と Hwang の *Introduction to Probability*（確率論入門）は、順序統計量と無記憶性をわかりやすく扱っている。

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。