解答例

STEP III 2003、第 12 問

確率・統計 ・ 指数分布に従う待ち時間

STEP III 2003、問 12

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/03-s3-q12

鍵となる考え方

任意の区間における中断の回数がポアソン分布（Poisson distribution）に従うならば、「次の中断まで t 分より長くかかる」という事象は「次の t 分間に中断が**一度も**起こらない」という事象と同じであり、これによって離散分布が連続分布に変わる。すなわち、待ち時間は指数分布（exponential distribution）に従う。さらに、中断と中断の間の間隔は次々に独立（independent）であるから、長い間隔を探すことは独立な試行（trials）の列となり、各試行は「成功」（30 分以上の間隔）か「失敗」のいずれかである。あとは幾何分布（geometric distribution）と条件付き期待値（conditional expectation）が片づけてくれる。

解答

T の分布

中断は平均して 10 分に 1 回起こるので、その率は 1 分あたり $\frac{1}{10}$ であり、t 分間の中断の回数は平均 $\frac{t}{10}$ のポアソン分布に従う。

たった今起こった中断の時点から時間を計り始める。次の中断までの時間 T が t 以上となるのは、次の t 分間に中断が起こらないときであり、かつそのときに限るから、

$$\mathrm{P}(T \geq t) = \mathrm{P}(\text{中断なし、時間}t) = \mathrm{e}^{-t/10}.$$

したがって、$t \geq 0$ に対して、作業時間が t 分未満となる確率は

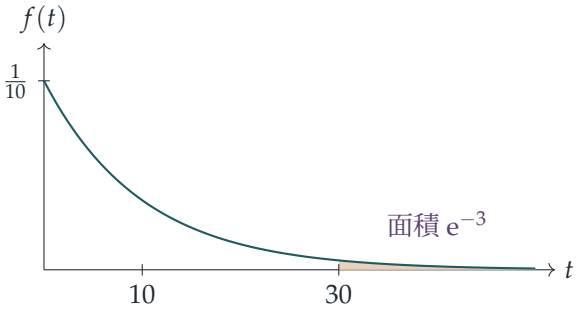
$$\mathrm{P}(T < t) = 1 - \mathrm{e}^{-t/10}.$$

これが T の累積分布関数（cumulative distribution function）$F(t)$ である（$t < 0$ に対

しては $F(t) = 0$)。微分すると

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

よって T は平均 10 分の指数分布に従う。



中断の回数の期待値

この過程には記憶がない（重なり合わない区間におけるポアソン分布の回数は互いに独立である）ので、中断が起こるたびに、次の中断までの時間はそれ以前のすべてとは独立に、T と同じ分布で新たに始まる。したがって、次々の間隔 $T_1, T_2, \ldots$ は独立で、それぞれ T と同じ分布に従う。

30 分以上の間隔を**長い**間隔とよぶことにする。各間隔は他とは独立に、確率

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

で長い間隔となる。最初の長い間隔より前に起こる中断の回数を N とする（たった今起こった中断は数えない）。短い間隔はどれも中断で終わり、最初の長い間隔が中断のない 30 分間となるので、N は最初の長い間隔より前の短い間隔の個数である。したがって

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

考えてみよう 1 ではなく 0 から始まるこの幾何分布の平均は、どのようにして求めればよいだろうか？

$|x| < 1$ のとき、幾何級数（geometric series）$\sum_{n\geq 0} x^n = \frac{1}{1-x}$ を微分すると $\sum_{n\geq 1} n x^{n-1} = \frac{1}{(1-x)^2}$ が得られる。$x = 1-p$ とおくと

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

（あるいは、$N+1$ は最初の成功までの（その成功を含む）試行回数であり、その平均

は $\frac{1}{p}$ である。）$p = \mathrm{e}^{-3}$ とすると

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

となり、求める結果が得られる。

短い間隔の長さの期待値

求めたいのは $\mathrm{E}(T \mid T < 30)$ である。$T < 30$ という条件のもとでの T の条件付き確率密度（density）は $0 \le t < 30$ において $f(t)/\mathrm{P}(T < 30)$ であるから、

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

$u = t$、$\mathrm{d}v = \frac{1}{10}\mathrm{e}^{-t/10}\,\mathrm{d}t$ とおいて部分積分（integration by parts）を行う。このとき $v = -\mathrm{e}^{-t/10}$ であるから、

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

したがって、分子と分母に e^3 を掛けると

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

これが 10 より少し小さいのは納得がいく。長い間隔を除外したからである。

待ち時間の期待値

中断のない 30 分間が始まるまでの待ち時間は、N 個の短い間隔の和である。$N = n$ という条件のもとでは、これらは n 個の独立な間隔で、それぞれ 30 分より短いという条件が課されているから、$N = n$ のもとでの待ち時間の期待値は $n\,\mathrm{E}(T \mid T < 30)$ である。N について平均をとると

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

となり、これは約 161 分、すなわち 2 時間 41 分である。（答案を**書き終える**までの時間を求めたいなら、30 分そのものを加えればよい：$10\, \mathrm{e}^3 - 10 \approx 191$ 分。）

検算

- $\int_0^{\infty} \frac{1}{10}\mathrm{e}^{-t/10}\,\mathrm{d}t = 1$ であるから f は確かに確率密度であり、その平均は 10 で、「平均して 10 分に 1 回」と一致する。

- 条件をつけない場合、$\mathrm{E}(T) = 10$ は $\mathrm{P}(T < 30)\,\mathrm{E}(T \mid T < 30) + \mathrm{P}(T \geq 30)\,\mathrm{E}(T \mid T \geq 30)$ と分解できるはずである。無記憶性（memorylessness）により $\mathrm{E}(T \mid T \geq 30) = 40$ であり、実際 $(10 - 40\mathrm{e}^{-3}) + 40\mathrm{e}^{-3} = 10$ となる。
- この過程を 200 000 回シミュレーションしたところ、中断の回数の平均は 19.1（$\mathrm{e}^3 - 1 \approx 19.09$ に対して）、待ち時間の平均は 161.1 分（$10\mathrm{e}^3 - 40 \approx 160.86$ に対して）であった。

よくある間違い

- $\mathrm{E}(N) = \frac{1}{p} = \mathrm{e}^3$ を用いること。これは長い間隔まで数えてしまっているが、長い間隔は 30 分間が始まる前に中断が起こらずに終わる。
- 待ち時間を求めるのに、$\mathrm{E}(N)$ に条件なしの平均 10 を掛けること。待ち時間を構成する間隔は短いことがわかっているので、その平均は 10 ではなく $\mathrm{E}(T \mid T < 30)$ である。

STEP のその先へ

中断は**ポアソン過程**（Poisson process）をなしており、間隔が独立な指数分布に従うという結果は、その定義的な性質の一つである。最後の段階 $\mathrm{E}(\text{全}N\text{ 項の和}) = \mathrm{E}(N) \times \mathrm{E}(\text{各項})$ は、**ワルドの等式**（Wald's identity）と密接に関連する条件付けの議論である。これらを合わせると、大学 1、2 年次の確率論の講義で学ぶ再生理論（renewal theory）や待ち行列モデル（queueing models）へとつながっていく。Grimmett と Welsh による *Probability: An Introduction*（確率論入門）は、ポアソン過程と条件付き期待値をちょうどよい水準で扱っている。

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。