解答例

STEP III 2021、第 11 問

確率・統計 ・ 指数分布

STEP III 2021、問 11

先に進む前に、まず問題を読んでおこう。問題は、OCR が配布している 2021 年の STEP 過去問一式（zip ファイル）のうち、STEP III の問題用紙にある：

www.ocr.org.uk/Images/696588-2021-past-papers.zip

この問題は、*Beyond STEP* の第 12.18 節「**無記憶性**」で詳しく解説されている：https://beyondstep.keithmcnulty.org/ja/ch12/memoryless.html

鍵となる考え方

Y と Z に関する問いはすべて、実は X がどの区間に入るかという問いである。事象 $Y = n$ は $n \leq X < n+1$ であり、事象 $Z < z$ は各 n に一つずつの、互いに排反（disjoint）な部分 $n \leq X < n+z$ に分かれる。指数分布（exponential distribution）の裾確率 $\mathrm{P}(X \geq x) = \mathrm{e}^{-\lambda x}$ によって、それぞれの部分は等比級数（geometric series）の項になる。

解答

以下を通して $q = \mathrm{e}^{-\lambda}$ と書く。$\lambda > 0$ であるから $0 < q < 1$ である。$x \geq 0$ に対して

$$\mathrm{P}(X \geq x) = \int_x^\infty \lambda \mathrm{e}^{-\lambda t}\,\mathrm{d}t = \mathrm{e}^{-\lambda x},$$

であり、X は連続型であるから、$0 \leq a < b$ に対して $\mathrm{P}(a \leq X < b) = \mathrm{P}(a < X < b) = \mathrm{e}^{-\lambda a} - \mathrm{e}^{-\lambda b}$ である。

小問 (i)

Y は X の整数部分（integer part）であるから、$Y = n$ となるのはちょうど $n \leq X < n+1$ のときである。よって

$$\mathrm{P}(Y = n) = \mathrm{e}^{-n\lambda} - \mathrm{e}^{-(n+1)\lambda} = (1 - \mathrm{e}^{-\lambda})\mathrm{e}^{-n\lambda},$$

となり、求める結果が得られる。（したがって Y は 0 から始まる公比 q の幾何分布（geometric distribution）に従う。）

小問 (ii)

$Z = X - Y$ は X の小数部分（fractional part）であるから、$0 \leq Z < 1$ である。$0 \leq z \leq 1$ とする。事象 $Z < z$ が起こるのは、整数 $n = Y$ について $n \leq X < n + z$ となるときであり、そのときに限る。$n = 0, 1, 2, \ldots$ に対するこれらの事象は互いに排反であるから

$$\begin{aligned} \mathrm{P}(Z < z) &= \sum_{n=0}^{\infty} \mathrm{P}(n \leq X < n + z) = \sum_{n=0}^{\infty} \left(\mathrm{e}^{-n\lambda} - \mathrm{e}^{-n\lambda - \lambda z}\right) \\ &= (1 - \mathrm{e}^{-\lambda z}) \sum_{n=0}^{\infty} q^n = \frac{1 - \mathrm{e}^{-\lambda z}}{1 - \mathrm{e}^{-\lambda}}, \end{aligned}$$

となる。ここで等比級数を和の公式で求めた。$0 < q < 1$ であるから、この級数は収束（converge）する。これが求める結果である。

小問 (iii)

(ii) の分布関数（distribution function）を微分すると、Z の確率密度関数（density）は

$$g(z) = \frac{\lambda \mathrm{e}^{-\lambda z}}{1 - \mathrm{e}^{-\lambda}} \qquad (0 \leq z \leq 1),$$

であり、それ以外では $g(z) = 0$ である。$u = z$、$\mathrm{d}v = \lambda \mathrm{e}^{-\lambda z}\,\mathrm{d}z$ として部分積分（integration by parts）を行うと

$$\int_0^1 z\lambda \mathrm{e}^{-\lambda z}\,\mathrm{d}z = \left[-z\mathrm{e}^{-\lambda z}\right]_0^1 + \int_0^1 \mathrm{e}^{-\lambda z}\,\mathrm{d}z = -\mathrm{e}^{-\lambda} + \frac{1 - \mathrm{e}^{-\lambda}}{\lambda}.$$

これを $1 - \mathrm{e}^{-\lambda}$ で割って

$$\mathrm{E}(Z) = \frac{1}{\lambda} - \frac{\mathrm{e}^{-\lambda}}{1 - \mathrm{e}^{-\lambda}} = \frac{1}{\lambda} - \frac{1}{\mathrm{e}^{\lambda} - 1}.$$

小問 (iv)

$0 \leq z_1 < z_2 \leq 1$ とする。事象 $\{Y = n \text{ かつ} z_1 < Z < z_2\}$ はちょうど $n + z_1 < X < n + z_2$ であるから

$$\mathrm{P}(Y = n \text{ かつ} z_1 < Z < z_2) = \mathrm{e}^{-\lambda(n+z_1)} - \mathrm{e}^{-\lambda(n+z_2)} = \mathrm{e}^{-n\lambda}\left(\mathrm{e}^{-\lambda z_1} - \mathrm{e}^{-\lambda z_2}\right).$$

? **考えてみよう** 独立性を判定するには、これを何と比べればよいだろうか？

積 $\mathrm{P}(Y=n)\,\mathrm{P}(z_1<\mathrm{Z}<z_2)$ と比べればよい。Z は連続型であるから $\mathrm{P}(\mathrm{Z}=z_1)=0$ であり、(ii) より

$$\mathrm{P}(z_1<\mathrm{Z}<z_2)=\mathrm{P}(\mathrm{Z}<z_2)-\mathrm{P}(\mathrm{Z}<z_1)=\frac{\mathrm{e}^{-\lambda z_1}-\mathrm{e}^{-\lambda z_2}}{1-\mathrm{e}^{-\lambda}}.$$

これに (i) の $\mathrm{P}(Y=n)=(1-\mathrm{e}^{-\lambda})\mathrm{e}^{-n\lambda}$ を掛けると、因数 $1-\mathrm{e}^{-\lambda}$ が約分され、上の同時確率がちょうど得られる。よって、すべての非負整数 n とすべての区間 $(z_1,z_2)\subseteq[0,1]$ に対して

$$\mathrm{P}(Y=n\text{ かつ}z_1<\mathrm{Z}<z_2)=\mathrm{P}(Y=n)\,\mathrm{P}(z_1<\mathrm{Z}<z_2)$$

が成り立つ。Y は非負整数値のみをとり、Z は $[0,1)$ の値のみをとるから、この形の事象によって同時分布（joint distribution）全体が決まる。したがって Y と Z は独立（independent）である。X がどの単位区間に入ったかを知っても、その区間のどこに入ったかについては何もわからないのである。

検算

- (ii) で $z=1$ とおくと、当然そうなるべきとおり $\mathrm{P}(\mathrm{Z}<1)=1$ となり、$z=0$ とおくと 0 となる。
- $\lambda\to0$ のとき、X の密度は各単位区間上でほぼ平坦になるから、Z は $[0,1]$ 上の一様分布（uniform distribution）に近くなるはずである。実際、$\mathrm{e}^{\lambda}-1=\lambda+\frac{1}{2}\lambda^2+\cdots$ より $\frac{1}{\mathrm{e}^{\lambda}-1}=\frac{1}{\lambda}-\frac{1}{2}+O(\lambda)$ となるので、$\mathrm{E}(Z)\to\frac{1}{2}$ である。$\lambda\to\infty$ のときは、X はほとんど常に非常に小さく、$\mathrm{E}(Z)\to0$ となる。
- $\lambda=1$ のとき $\mathrm{E}(Z)=1-\frac{1}{\mathrm{e}-1}\approx0.4180$ であり、これは X の値 200 000 個によるコンピュータシミュレーションと一致した。(iv) のシミュレーションによる同時度数も、周辺分布（marginal distribution）の積と一致した。

よくある間違い

- $\mathrm{P}(\mathrm{Z}<z)=\mathrm{P}(X<z)$ と書いてしまうこと。$\mathrm{Z}<z$ は最初の単位区間だけでなく、すべての単位区間で起こりうることを忘れている。
- n、z_1、z_2 を一組だけ選んで独立性を結論づけること。積への分解はそれらすべてについて成り立たなければならない。

STEP のその先へ

(iv) の独立性は、指数分布の**無記憶性**（memoryless property）の現れである。すなわち、$X \geq n$ という条件のもとで、超過分 $X - n$ は n が何であっても再び同じパラメータの指数分布に従う。これこそ、指数分布がポアソン過程（Poisson process）の待ち時間のモデルとなる理由であり、またその整数部分が離散型の無記憶な分布である幾何分布に従う理由でもある。Grimmett and Welsh, *Probability: An Introduction*（確率論入門）では、これら二つの分布とポアソン過程が詳しく扱われている。

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。