解答例

STEP III 2019、第 11 問

確率・統計 ・ ポアソン間引き

STEP III 2019、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 2019 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/696590-2019-past-papers.zip

鍵となる考え方

すべては、やって来る客の人数、あるいは砂を取る客の人数で条件付け（conditioning）を行い、そのうえでポアソン型の級数（series）の和を求めることに帰着する。小問 (i) はポアソン分布（Poisson distribution）の古典的な「間引き」（thinning）である。二項分布（binomial distribution）の確率とポアソン分布の確率を掛け合わせると、観測されない総数についての和が指数関数にまとまる。砂を取る客の人数がポアソン分布に従うとわかれば、小問 (ii) と (iii) で必要なのは $(1-k)^Y$ の期待値（expectation）だけであり、これもまた指数関数の級数になる。

解答

1 日の客の人数を N とすると $N \sim \mathrm{Po}(\lambda)$ であり、そのうち砂を取る客の人数を Y とする。各客は互いに独立に、また N とも独立に判断すると仮定する。したがって $N = n$ のもとで $Y \sim \mathrm{B}(n, p)$ である。

小問 (i)

$y = 0, 1, 2, \ldots$ に対して、N で条件付けると（寄与するのは $n \geq y$ のみである）、

$$\begin{aligned} \mathrm{P}(Y = y) &= \sum_{n=y}^{\infty} \mathrm{e}^{-\lambda} \frac{\lambda^n}{n!} \binom{n}{y} p^y (1-p)^{n-y} \\ &= \mathrm{e}^{-\lambda} \frac{(p\lambda)^y}{y!} \sum_{n=y}^{\infty} \frac{(\lambda(1-p))^{n-y}}{(n-y)!} \qquad \text{なぜなら} \frac{\lambda^n}{n!} \cdot \frac{n!}{y!\,(n-y)!} = \frac{\lambda^y \lambda^{n-y}}{y!\,(n-y)!} \\ &= \mathrm{e}^{-\lambda} \frac{(p\lambda)^y}{y!} \mathrm{e}^{\lambda(1-p)} = \mathrm{e}^{-p\lambda} \frac{(p\lambda)^y}{y!}, \end{aligned}$$

となる。ここで $m = n - y$ とおき、$\sum_{m \geq 0} x^m / m! = \mathrm{e}^x$ を用いた。これは $\mathrm{Po}(p\lambda)$ の確率であるから、求めるとおり $Y \sim \mathrm{Po}(p\lambda)$ である。

小問 (ii)

砂を取る客は、それぞれその時点で残っている量の $1 - k$ の割合を残していく。よって Y 人が砂を取った後に残る質量は $S(1-k)^Y$ であり、取られた質量は

$$T = S(1 - (1-k)^Y).$$

である。小問 (i) より

$$\mathrm{E}[(1-k)^Y] = \sum_{y=0}^{\infty} (1-k)^y \mathrm{e}^{-p\lambda} \frac{(p\lambda)^y}{y!} = \mathrm{e}^{-p\lambda} \mathrm{e}^{p\lambda(1-k)} = \mathrm{e}^{-kp\lambda}.$$

したがって、期待値の線形性（linearity of expectation）により

$$\mathrm{E}(T) = S(1 - \mathrm{E}[(1-k)^Y]) = (1 - \mathrm{e}^{-kp\lambda})S,$$

となり、これが求める結果である。

小問 (iii)

> **考えてみよう** 「砂の k の割合」を、特定の 1 粒についての確率にどう言い換えればよいだろうか？

砂粒は非常に多く、どの粒も特別扱いされないので、誰かが残りの砂の k の割合を取るたびに、まだ袋の中にある粒は、それ以前に起きたこととは独立に確率 k で取られる。（言い換えれば、金の粒がある取り分に含まれる確率は、その取り分が全質量に占める割合に等しい。）

助手が金の粒を取るという事象を G とする。$Y = y$ のもとで、金の粒が y 人の客すべてを生き残る確率は $(1-k)^y$ であり、その後に助手がそれを取る確率は k である。よ

って

$$\mathrm{P}(G \mid Y = y) = k(1-k)^y,$$

であり、小問 (ii) で求めた期待値を用いると

$$\mathrm{P}(G) = k\,\mathrm{E}[(1-k)^Y] = k\,\mathrm{e}^{-kp\lambda}.$$

となる。これは小問 (ii) と整合している。金の粒がその日を生き残る確率は $\mathrm{e}^{-kp\lambda}$ であり、これはちょうど残る砂の割合の期待値に等しい。

$k = 0$ の場合。 このとき $\mathrm{P}(G) = 0$ である。これは納得がいく。$k = 0$ ならば助手も含めて誰も砂を取らないので、金の粒は袋の中に残る。

極限 $k \to 1$。 このとき $\mathrm{P}(G) \to \mathrm{e}^{-p\lambda}$ であり、これは $\mathrm{P}(Y = 0)$ に等しい。k が 1 に近いとき、最初に砂を取る客がほぼすべての砂を取り、金の粒もほぼ確実にそこに含まれる。助手が金の粒を得るのは、実質的にどの客も砂を取らなかった場合に限られ、そのとき助手は（ほぼ）袋の砂すべてを取る。

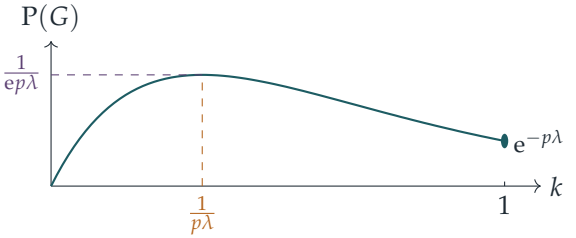
最大化。 $\mu = p\lambda$ とおき、$0 \le k < 1$ に対して $f(k) = k\,\mathrm{e}^{-\mu k}$ とする。すると

$$f'(k) = \mathrm{e}^{-\mu k}(1 - \mu k),$$

であり、これは $k < 1/\mu$ で正、$k > 1/\mu$ で負である。$\mu > 1$ であるから点 $k = 1/\mu$ は $(0,1)$ に含まれ、f はそこまで増加し、その後は減少する。よってこの点で $[0,1)$ 上の最大値をとる。したがって確率が最大となるのは

$$k = \frac{1}{p\lambda},$$

のときであり、最大の確率は $\dfrac{1}{\mathrm{e}\,p\lambda}$ である。（$p\lambda \le 1$ ならば f は $[0,1)$ 全体で増加し、最大値は存在せず、$k \to 1$ のときの上限（supremum）$\mathrm{e}^{-p\lambda}$ があるだけである。問題が $p\lambda > 1$ に限定しているのはこのためである。）



グラフは $p\lambda = 3$ の場合の $\mathrm{P}(G) = k\,\mathrm{e}^{-kp\lambda}$ を示している。

検算

- 小問 (i) は明らかな確認をすべて通る。$p = 1$ とすれば $\mathrm{Po}(\lambda)$ に戻り、$p = 0$ とすれば確実に $Y = 0$ となる。
- 小問 (ii) では、$k = 0$ とすれば取られる砂は 0 であり、$\lambda \to \infty$ のとき取られる質量の期待値は S に近づく。いずれも当然そうなるべき結果である。
- $\lambda = 3$、$p = 0.6$、$k = 0.4$ のとき、公式は $\mathrm{P}(G) = 0.4\,\mathrm{e}^{-0.72} \approx 0.1947$ を与える。200 000 日分のコンピュータシミュレーションでは 0.1953 となった。
- $p\lambda = 3$ のとき、細かい格子上で $k\,\mathrm{e}^{-3k}$ が最大となるのは $k \approx 0.3333$ であり、その値は $0.1226 \approx 1/(3\mathrm{e})$ である。

よくある間違い

- 各客が**もとの**砂の k の割合を取るかのように、取られた砂を kSY と書いてしまう。各客は**残っている**砂の一部を取るので、取られる量は等比数列（geometric sequence）をなす。
- $\mathrm{E}[(1-k)^Y]$ を $(1-k)^{\mathrm{E}(Y)} = (1-k)^{p\lambda}$ と計算してしまう。非線形な関数の期待値は、期待値の関数ではない。
- $1/(p\lambda)$ が許される範囲 $0 \le k < 1$ に実際に入っているかの確認を忘れる。条件 $p\lambda > 1$ が使われるのはまさにここである。

STEP のその先へ

小問 (i) はポアソン分布の間引きの性質である。小問 (ii) と (iii) に現れる期待値 $\mathrm{E}[s^Y] = \mathrm{e}^{\mu(s-1)}$ は、$\mathrm{Po}(\mu)$ の**確率母関数**（probability generating function）を $s = 1-k$ で評価したものである。母関数を使えば、間引きはほぼ 1 行で示せる。$\mathrm{E}[s^N] = \mathrm{e}^{\lambda(s-1)}$ ならば $\mathrm{E}[s^Y] = \mathrm{E}[(1-p+ps)^N] = \mathrm{e}^{\lambda p(s-1)}$ である。ポアソン過程（Poisson process）の理論では、これはポアソン過程の点に独立に印を付けると再びポアソン過程が得られる、という主張になる。Grimmett and Welsh, *Probability: An Introduction*（確率論入門）は、母関数とこれらの考え方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。