解答例

STEP I 1995、第 13 問

確率・統計 ・ ポアソン間引き

> **STEP I 1995、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/95-s1-q13

鍵となる考え方

小問 (i) はポアソン分布（Poisson distribution）の**間引き**（thinning）の性質である。がん細胞の真の個数で条件づけると、印をつけられた細胞の個数は二項分布（binomial distribution）に従う。未知の真の個数について和をとると、きれいに指数級数（exponential series）にまとまる。小問 (ii) では、「2 個目の印がスライド k につく」という事象は、最初の $k-1$ 枚のスライドにつけられた印の総数（0 個か 1 個か）によって決まる。そして小問 (i) と、独立なポアソン変数の加法性（additivity）とから、その総数がどのように分布するかが正確にわかる。

解答

以下を通して $q = 1 - p$、$\lambda = \mu p$ とおく。異なるスライド上のがん細胞の個数は互いに独立（independent）であるとし（「無作為」の自然な解釈である）、科学者による検出は互いに独立であり、かつ細胞の個数とも独立であるとする。

小問 (i)

1 枚のスライドについて、がん細胞の個数を N、彼女が印をつける個数を M とすると、$N \sim \mathrm{Po}(\mu)$ である。$N = n$ のもとで、n 個の細胞はそれぞれ独立に確率 p で発見されるので、$M \mid N = n \sim \mathrm{B}(n, p)$ となる。$m = 0, 1, 2, \ldots$ に対して、全確率の法則（law of

total probability）により（寄与するのは $n \geq m$ のみである）、

$$\begin{aligned} \mathrm{P}(M=m) &= \sum_{n=m}^{\infty} \mathrm{e}^{-\mu}\frac{\mu^n}{n!}\binom{n}{m}p^m q^{n-m} = \mathrm{e}^{-\mu}\frac{(\mu p)^m}{m!}\sum_{n=m}^{\infty}\frac{(\mu q)^{n-m}}{(n-m)!} \\ &= \mathrm{e}^{-\mu}\frac{(\mu p)^m}{m!}\mathrm{e}^{\mu q} = \mathrm{e}^{-\mu p}\frac{(\mu p)^m}{m!}, \end{aligned}$$

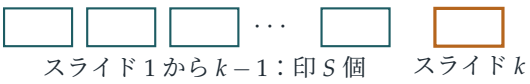
ここで $\binom{n}{m}/n! = 1/(m!\,(n-m)!)$ を用い、次に $j = n-m$ として指数級数を用い、最後に $-\mu+\mu q = -\mu p$ を用いた。よって、求めるとおり $M \sim \mathrm{Po}(\mu p)$ である。

小問 (ii)

スライド i 上で印をつけられた細胞の個数を M_i とする。小問 (i) とスライド間の独立性により、$M_1, M_2, \ldots$ は独立な $\mathrm{Po}(\lambda)$ 変数である。最初の $k-1$ 枚のスライドで印をつけられた個数を $S = M_1 + \cdots + M_{k-1}$ とする。独立なポアソン変数の和は、平均を足し合わせたポアソン分布に従うので $S \sim \mathrm{Po}((k-1)\lambda)$ であり、S は M_k と独立である。

> **考えてみよう** 2 個目の印がスライド k につくためには、正確には何が起こらなければならないか？

スライド k より前には印が高々 1 個で、スライド k の終わりまでには少なくとも 2 個でなければならない。したがって、互いに排反な 2 つの場合がある：



$S = 0$ かつ $M_k \geq 2$、または $S = 1$ かつ $M_k \geq 1$

（$k = 1$ のときはそれ以前のスライドはなく、自動的に $S = 0$ となり、以下の式はそのまま成り立つ。）S と M_k の独立性を用いると、

$$\begin{aligned} \mathrm{P}(S=0,\ M_k \geq 2) &= \mathrm{e}^{-(k-1)\lambda}\left(1-\mathrm{e}^{-\lambda}-\lambda\mathrm{e}^{-\lambda}\right), \\ \mathrm{P}(S=1,\ M_k \geq 1) &= (k-1)\lambda\,\mathrm{e}^{-(k-1)\lambda}\left(1-\mathrm{e}^{-\lambda}\right). \end{aligned}$$

これらを足して、

$$Q = \mathrm{e}^{-(k-1)\lambda}\left\{\left(1+(k-1)\lambda\right)\left(1-\mathrm{e}^{-\lambda}\right)-\lambda\mathrm{e}^{-\lambda}\right\}.$$

求める形に合わせるため、$1+(k-1)\lambda = (1+k\lambda)-\lambda$ と書くと、

$$\begin{aligned} \left(1+(k-1)\lambda\right)\left(1-\mathrm{e}^{-\lambda}\right)-\lambda\mathrm{e}^{-\lambda} &= (1+k\lambda)\left(1-\mathrm{e}^{-\lambda}\right)-\lambda+\lambda\mathrm{e}^{-\lambda}-\lambda\mathrm{e}^{-\lambda} \\ &= (1+k\lambda)\left(1-\mathrm{e}^{-\lambda}\right)-\lambda. \end{aligned}$$

$\lambda = \mu p$ とすると

$$Q = \mathrm{e}^{-\mu p(k-1)}\left\{(1+k\mu p)(1-\mathrm{e}^{-\mu p}) - \mu p\right\},$$

となり、これが示すべき結果である。

検算

- $k = 1$ のとき、式は $1 - \mathrm{e}^{-\lambda} - \lambda\mathrm{e}^{-\lambda} = \mathrm{P}(M_1 \geq 2)$ を与え、当然そうなるべき値になっている。
- 確率を $k = 1, 2, \ldots$ にわたって足すと 1 になるはずである（彼女はいずれ確率 1 で 2 個目の細胞に印をつける）。いくつかの λ の値について数値的に和をとると、このことが確かめられる。
- $\mu = 1.6$、$p = 0.45$（したがって $\lambda = 0.72$）でシミュレーションを行うと、200 000 回の試行のうち約 26.0% で 2 個目の印がスライド 2 についた。式からは $Q \approx 0.2591$ が得られる。

よくある間違い

- 「スライド k より前に印が 1 個、スライド k に少なくとも 1 個」という場合を忘れる、あるいは誤って $S = 1$、$M_k = 0$ の場合を含めてしまう。
- 「最初に印のついたスライド」といった議論でスライドごとに考えると、煩雑になる。最初の $k-1$ 枚のスライドにわたる総数 S を考えるのが効率的である。
- 独立性（スライド間の独立性、および S と M_k の独立性）をどこで用いたかを述べない。

STEP のその先へ

間引きはポアソン過程（Poisson process）の一つの側面である。事象がポアソン過程に従って起こり、各事象を独立に確率 p で残すとき、残された事象もまたポアソン過程をなし、その強度は p 倍になる。すると小問 (ii) は、2 番目の事象までの待ち時間（waiting time）の離散版であり、連続時間ではこれはガンマ分布（gamma distribution）に従う。Grimmett and Welsh, *Probability: An Introduction*（確率論入門）では、ポアソン過程とこうした待ち時間の分布が展開されている。

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。