解答例

STEP III 1992、第 16 問

確率・統計 ・ ポアソン間引き

> **STEP III 1992、問 16**
>
> 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/92-s3-q16

鍵となる考え方

誤植の総数を固定すれば、各誤植は私が見つけるか見逃すかのどちらかである独立な試行となるので、小問 (i) は二項分布（binomial distribution）の確率である。小問 (ii) は条件付き確率（conditional probability）であり、その分母、すなわちちょうど m 個の誤植を見つける確率を求めるには、見逃した誤植の個数としてありうるすべての値について和をとる必要がある。この和は指数関数の級数（exponential series）になることがわかり、そのおかげですべてがきれいなポアソン分布（Poisson distribution）の形にまとまるのである。

解答

以下では $q = 1 - p$ と書く。その号に含まれる誤植の総数を N、私が見つける誤植の個数を S、見逃す誤植の個数を $U = N - S$ とする。問題文より

$$\mathrm{P}(N = k) = \frac{\mathrm{e}^{-\lambda}\lambda^k}{k!}, \qquad k = 0, 1, 2, \ldots$$

である。

小問 (i)

$N = m + n$ のもとで、$m + n$ 個の誤植はそれぞれ他と独立に確率 p で見つけられる。したがって、$N = m + n$ という条件のもとで、見つける誤植の個数は二項分布 $\mathrm{B}(m + n, p)$ に従う。見つける m 個と見逃す n 個の特定の選び方 1 つの確率は $p^m q^n$ で

あり、そのような選び方は $\binom{m+n}{m}$ 通りあるので、

$$\mathrm{P}(S = m \mid N = m+n) = \binom{m+n}{m} p^m (1-p)^n.$$

小問 (ii)

求めたいのは $\mathrm{P}(U = n \mid S = m)$ である。事象 $\{S = m,\, U = n\}$ は $\{N = m+n,\, S = m\}$ と同じであるから、小問 (i) より

$$\begin{aligned}\mathrm{P}(S = m,\, U = n) &= \mathrm{P}(N = m+n)\,\mathrm{P}(S = m \mid N = m+n) \\ &= \frac{\mathrm{e}^{-\lambda}\lambda^{m+n}}{(m+n)!} \cdot \frac{(m+n)!}{m!\,n!}\, p^m q^n \\ &= \mathrm{e}^{-\lambda}\,\frac{(\lambda p)^m}{m!} \cdot \frac{(\lambda q)^n}{n!}.\end{aligned}$$

因子 $(m+n)!$ が約分されて消えることが、この問題の要である。

考えてみよう 誤植が何個あったかわからないとき、$\mathrm{P}(S = m)$ はどのように求めればよいだろうか。

ちょうど m 個を見つけたとき、見逃した個数は $0, 1, 2, \ldots$ のどれでもありうり、これらの場合は互いに排反である。そこで、同時確率を U のすべての値 j について足し合わせる：

$$\begin{aligned}\mathrm{P}(S = m) &= \sum_{j=0}^{\infty} \mathrm{e}^{-\lambda}\,\frac{(\lambda p)^m}{m!} \cdot \frac{(\lambda q)^j}{j!} = \mathrm{e}^{-\lambda}\,\frac{(\lambda p)^m}{m!}\,\mathrm{e}^{\lambda q} \\ &= \frac{(\lambda p)^m}{m!}\,\mathrm{e}^{-\lambda p},\end{aligned}$$

ここで $\sum_{j\geq 0} x^j/j! = \mathrm{e}^x$ と $-\lambda + \lambda q = -\lambda p$ を用いた。（したがって、私が見つける誤植の個数そのものも平均 λp のポアソン分布に従う。）これは正であるから、これで割ってよい。よって

$$\begin{aligned}\mathrm{P}(U = n \mid S = m) &= \frac{\mathrm{P}(S = m,\, U = n)}{\mathrm{P}(S = m)} = \frac{\mathrm{e}^{-\lambda}\,\dfrac{(\lambda p)^m}{m!} \cdot \dfrac{(\lambda q)^n}{n!}}{\mathrm{e}^{-\lambda p}\,\dfrac{(\lambda p)^m}{m!}} \\ &= \frac{(\lambda q)^n}{n!}\,\mathrm{e}^{-\lambda+\lambda p} = \frac{(1-p)^n\lambda^n}{n!}\,\mathrm{e}^{-(1-p)\lambda},\end{aligned}$$

となり、示された。

この答えは m にまったく依存しない。すなわち、私が誤植をいくつ見つけたかを知っても、いくつ見逃したかについては何の情報も得られない。見逃した個数は、m が何であっても平均 $(1-p)\lambda$ のポアソン分布に従うのである。

検算

- 最終的な答えを $n = 0, 1, 2, \ldots$ について足し合わせると $\mathrm{e}^{(1-p)\lambda}\mathrm{e}^{-(1-p)\lambda} = 1$ となり、条件付き分布として当然満たすべき性質を満たしている。
- $p \to 1$ のときは私がすべてを見つけることになり、答えは $n = 0$ のとき 1 に、$n \geq 1$ のとき 0 に収束する。妥当である。
- $\lambda = 3$、$p = 0.6$、$m = 2$、$n = 1$ のとき、公式は $1.2\,\mathrm{e}^{-1.2} \approx 0.3614$ を与える。400 000 部の新聞をシミュレーションし、ちょうど 2 個の誤植が見つかったものだけを残したところ、0.360 という値が得られた。

よくある間違い

- 小問 (i) で二項係数 $\binom{m+n}{m}$ を書き忘れること。見つけた m 個の誤植は、$m+n$ 個のうちのどの m 個でもありうる。
- 見逃した誤植の個数としてありうるすべての値について和をとらずに、$\mathrm{P}(S = m)$ を $\mathrm{P}(N = m)$ や小問 (i) の答えとしてしまうこと。

STEP のその先へ

これはポアソン分布の**間引き**（thinning、または分割 splitting）と呼ばれる性質である。ポアソン (λ) 分布に従う個数の対象それぞれを確率 p で残すとき、残った個数と捨てた個数は、それぞれ平均 λp と $\lambda(1-p)$ をもつ**独立な**ポアソン確率変数となる。今回求めた同時確率は、まさにこの 2 つの部分の積に分解されている。間引きはポアソン過程（Poisson process）の理論における重要な道具であり、ランダムな事象の 1 つの流れを独立な部分的な流れに分割することを可能にする。Grimmett と Welsh による *Probability: An Introduction*（確率論入門）はこの性質を扱っており、それを 1 行で証明する母関数（generating function）の方法もあわせて解説している。

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。