解答例

STEP II 2005、第 13 問

確率・統計 ・ ポアソン分布と二項分布

> **STEP II 2005、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/05-s2-q13

鍵となる考え方

抽出した各ページについて、誤りが 2 個未満であるか否かのどちらかであり、各ページは互いに独立（independent）にふるまう。したがって Y は n 回の独立な試行における成功の回数を数えており、二項分布（binomial distribution）に従う。それ以降はすべて λ が小さいときに何が起こるかという問題であり、そのための道具は $\mathrm{e}^{-\lambda}$（あるいは $\ln(1+\lambda)$）のマクローリン展開（Maclaurin series）である。打ち消し合わずに残る最初の項だけを残せばよい。

解答

Y の分布

ある特定のページ上の誤りの個数を X とすると、$X \sim \mathrm{Po}(\lambda)$ である。$X \leq 1$ のとき、そのページを**良い**ページとよぶことにする。このとき

$$\mathrm{P}(\text{良い}) = \mathrm{P}(X=0) + \mathrm{P}(X=1) = \mathrm{e}^{-\lambda} + \lambda\mathrm{e}^{-\lambda} = (1+\lambda)\mathrm{e}^{-\lambda} = p.$$

標本は n 枚の異なるページからなる。異なるページ上の誤りの個数は互いに独立で、いずれも同じポアソン分布（Poisson distribution）$\mathrm{Po}(\lambda)$ に従う。よって、どの n ページを選んでも、事象「ページ j は良い」は互いに独立であり、それぞれの確率は p である。（条件 $n \ll N$ は、抽出元のページが十分たくさんあることを保証しているにすぎない。独立性はモデルから得られるものであり、この条件は独立性には必要ない。）

したがって Y は、成功確率 p の独立な試行を n 回行ったときの成功回数である。特定の k 枚の良いページと $n-k$ 枚の悪いページの組が現れる確率は $p^k q^{n-k}$ であり、どの

k ページが良いかの選び方は $\binom{n}{k}$ 通りあるから、

$$\mathrm{P}(Y=k)=\binom{n}{k}p^k q^{n-k},\qquad k=0,1,\ldots,n,$$

ただし $p=(1+\lambda)\mathrm{e}^{-\lambda}$、$q=1-p$ であり、これが示すべきことであった。すなわち $Y\sim \mathrm{B}(n,p)$ である。

小問 (i)

$\mathrm{e}^{-\lambda}=1-\lambda+\frac{1}{2}\lambda^2-\frac{1}{6}\lambda^3+\cdots$ と展開し、$1+\lambda$ を掛けると

$$\begin{aligned}p=(1+\lambda)\mathrm{e}^{-\lambda}&=(1-\lambda+\tfrac{1}{2}\lambda^2-\tfrac{1}{6}\lambda^3+\cdots)+(\lambda-\lambda^2+\tfrac{1}{2}\lambda^3-\cdots)\\&=1-\tfrac{1}{2}\lambda^2+\tfrac{1}{3}\lambda^3-\cdots.\end{aligned}$$

λ の項が打ち消し合うので、λ が小さいとき

$$q=1-p=\tfrac{1}{2}\lambda^2-\tfrac{1}{3}\lambda^3+\cdots\approx\tfrac{1}{2}\lambda^2$$

となる。無視した項は $\frac{1}{2}\lambda^2$ に比べて λ 程度の因子だけ小さいからである。

考えてみよう $\mathrm{e}^{-\lambda}$ の最初の項から想像されるように $q\approx\lambda$ とならないのはなぜか？

$\mathrm{e}^{-\lambda}$ から出る $-\lambda$ が、$1+\lambda$ を掛けたことで生じる $+\lambda$ とちょうど打ち消し合うからである。ページが悪いのは誤りが 2 個以上あるときに限り、$\mathrm{P}(X\geq 2)\approx\mathrm{P}(X=2)=\frac{1}{2}\lambda^2\mathrm{e}^{-\lambda}\approx\frac{1}{2}\lambda^2$ である。これは (i) を別の方法で確かめるよい検証になっている。

小問 (ii)

$\mathrm{P}(Y=n)=p^n$ であり、$p^n\geq 1-\lambda$ を満たす最大の n を求めればよい。（$0<\lambda<1$ のとき）両辺は正であり、ln は増加関数なので、これは

$$n\ln p\geq\ln(1-\lambda).$$

と同値である。ここで $\ln p=\ln(1+\lambda)-\lambda=(\lambda-\frac{1}{2}\lambda^2+\cdots)-\lambda\approx-\frac{1}{2}\lambda^2$ であり、$\ln(1-\lambda)\approx-\lambda$ である。$\ln p<0$ なので、これで割ると不等号の向きが逆になる：

$$n\leq\frac{\ln(1-\lambda)}{\ln p}\approx\frac{-\lambda}{-\frac{1}{2}\lambda^2}=\frac{2}{\lambda}.$$

よって、そのような最大の n はおよそ $2/\lambda$ である。

(i) を用いて同じことを示すこともできる：$p^n=(1-q)^n\approx 1-nq\approx 1-\frac{1}{2}n\lambda^2$ であり、$1-\frac{1}{2}n\lambda^2\geq 1-\lambda$ となるのはちょうど $n\leq 2/\lambda$ のときである。$(1-q)^n\approx 1-nq$ という近似には nq が小さいことが必要だが、ここでは $nq\approx\frac{2}{\lambda}\cdot\frac{1}{2}\lambda^2=\lambda$ であり小さいので、この近似は自己無撞着である。

小問 (iii)

$Y > 1$ ならば $Y > 0$ であるから、条件付き確率（conditional probability）は

$$\mathrm{P}(Y > 1 \mid Y > 0) = \frac{\mathrm{P}(Y > 1)}{\mathrm{P}(Y > 0)} = \frac{\mathrm{P}(Y > 0) - \mathrm{P}(Y = 1)}{\mathrm{P}(Y > 0)} = 1 - \frac{npq^{n-1}}{1 - q^n}.$$

（ここで $n \geq 2$ が効いている。これにより $Y > 1$ は起こりうる事象となり、べき q^{n-1} は少なくとも q となる。）次に n を固定して λ を小さくする。このとき $q \approx \frac{1}{2}\lambda^2$ は小さいので、$p \approx 1$ かつ $1 - q^n \approx 1$ となり、

$$\mathrm{P}(Y > 1 \mid Y > 0) \approx 1 - nq^{n-1} \approx 1 - n\left(\frac{\lambda^2}{2}\right)^{n-1},$$

を得る。最後の段階では (i) を用いた。これが求める結果である。

検算

- $\lambda = 0.8$ として、$n = 5$ ページのポアソン標本を 200,000 回シミュレーションしたところ、$Y = 0, \ldots, 5$ の相対度数は $\mathrm{B}(5, p)$ と 0.002 以内で一致した。
- $\lambda = 0.01$ のとき、$q = 4.967 \times 10^{-5}$ であり、$\frac{1}{2}\lambda^2 = 5 \times 10^{-5}$ と比べてよく合う。
- $\lambda = 0.01$ のとき、$p^n \geq 0.99$ を満たす最大の n の正確な値は 202 であり、$2/\lambda = 200$ と比べてよく合う。（各級数でもう 1 項多く残すと $2/\lambda + \frac{7}{3}$ となるので、$2/\lambda$ が大きくなっても誤差は有界にとどまる。）
- $\lambda = 0.01$、$n = 3$ のとき、正確な値は $1 - \mathrm{P}(Y > 1 \mid Y > 0) = 7.40 \times 10^{-9}$ であり、$3(\lambda^2/2)^2 = 7.5 \times 10^{-9}$ と比べてよく合う。

よくある間違い

- 「誤りが 2 個未満」を $X = 0$ のみ、あるいは $X \leq 2$ と解釈してしまう。
- $(1 + \lambda)\mathrm{e}^{-\lambda}$ の展開を 1 次の項で打ち切り、$q \approx 0$ と結論してしまう。残る最初の項は 2 次の項である。
- (ii) で $\ln p < 0$ で割るときに、不等号の向きを逆にし忘れる。

STEP のその先へ

小問 (ii) は漸近解析（asymptotic analysis）の一例である。条件は積 $n\lambda^2$ によって支配されるので、n の自然なスケールは $1/\lambda$ であり、p^n のような量は $\mathrm{e}^{-n\lambda^2/2}$ のようにふ

るまう。同じ考え方が、二項分布のポアソン近似や、確率論における多くの「まれな事象」の極限の根底にある。Grimmett と Welsh による *Probability: An Introduction*（確率論入門）は、これらの極限定理を丁寧に展開している。

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。