解答例

STEP I 2010、第 12 問

確率・統計 ・ 裾の和を用いた期待値

STEP I 2010、問 12

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/10-s1-q12

鍵となる考え方

裾和公式（tail-sum formula）は、$n\,\mathrm{P}(X=n)$ を $\mathrm{P}(X=n)$ を n 回足したものとして書き、項を別の順序でまとめ直すことから得られる。ペンギンの問題では、事象 $X \geq n$ のほうが $X=n$ よりはるかに記述しやすい。$n-1$ 個の箱を開けたあとでさらに箱を開ける必要があるのは、ちょうどその $n-1$ 個の箱にすべて同じ種類のペンギンが入っているときである。したがって裾確率は簡単な形になり、いま証明した公式によって $\mathrm{E}(X)$ は 2 つの等比級数（geometric series）になる。

解答

定義と裾和公式

X は $1, 2, 3, \ldots$ の値しかとらないので、

$$\mathrm{E}(X) = \sum_{m=1}^{\infty} m\,\mathrm{P}(X=m),$$

である。ただし、この級数が収束（converge）する場合に限る。

各 $n \geq 1$ について、事象 $X \geq n$ は $m \geq n$ に対する事象 $X = m$ の互いに排反な

（disjoint）和事象であるから、$\mathrm{P}(X \geq n) = \sum_{m \geq n} \mathrm{P}(X = m)$ である。よって

$$\begin{aligned}\sum_{n=1}^{\infty} \mathrm{P}(X \geq n) &= \sum_{n=1}^{\infty} \sum_{m=n}^{\infty} \mathrm{P}(X = m) \\ &= \sum_{m=1}^{\infty} \sum_{n=1}^{m} \mathrm{P}(X = m) \\ &= \sum_{m=1}^{\infty} m\,\mathrm{P}(X = m) = \mathrm{E}(X).\end{aligned}$$

考えてみよう なぜ 2 つの和の順序を入れ替えてよいのだろうか。

この二重和は、$1 \leq n \leq m$ を満たすすべての組 (n, m) について $\mathrm{P}(X = m)$ を足し合わせたものである。1 行目では n を固定して m について和をとり、2 行目では m を固定して $n = 1, \ldots, m$ の m 個の値について和をとっている。どの項も非負であり、非負の項からなる級数では、総和は和をとる順序によらない（両辺が無限大になることもあり、その場合は両辺とも無限大である）。項を三角形状に並べ、第 m 行に $\mathrm{P}(X = m)$ を m 個置いた図を思い描くとよい。行ごとに足せば $\mathrm{E}(X)$ が得られ、列ごとに足せば裾和が得られる。これで公式が証明された。

裾確率

意図されているとおり、異なる箱の中身は独立（independent）であると仮定する。D_k を箱 k に父ペンギンが入っているという事象とすると、$\mathrm{P}(D_k) = p$、$\mathrm{P}(D_k') = q$ である。

$n \geq 2$ とする。$X \geq n$ となるのは、最初の $n-1$ 個の箱にまだ両方の種類がそろっていないとき、すなわちすべて父ペンギンであるか、すべて母ペンギンであるときであり、かつそのときに限る。この 2 つの場合は互いに排反であるから、独立性より

$$\mathrm{P}(X \geq n) = p^{n-1} + q^{n-1} \qquad (n \geq 2).$$

特に $n = 4$ とすると

$$\mathrm{P}(X \geq 4) = p^3 + q^3,$$

となり、求める結果が得られる。また、1 つの箱に両方の種類が入ることはないので常に $X \geq 2$ であり、$\mathrm{P}(X \geq 1) = \mathrm{P}(X \geq 2) = 1$ である。（上の式は $n = 2$ では $p + q = 1$ を与えるので一致するが、$n = 1$ では 2 を与えてしまうので、$n = 1$ の項は別に扱わなければならない。）

期待値

裾和公式より

$$\begin{aligned}\mathrm{E}(X) &= \mathrm{P}(X \geq 1) + \sum_{n=2}^{\infty} (p^{n-1} + q^{n-1}) \\ &= 1 + \sum_{k=1}^{\infty} p^k + \sum_{k=1}^{\infty} q^k \\ &= 1 + \frac{p}{1-p} + \frac{q}{1-q},\end{aligned}$$

である。ここで、$0 < p < 1$ かつ $0 < q < 1$ であるから、どちらの等比級数も収束する（$p \neq 0$、$q \neq 0$ という条件はここで使われる）。$1 - p = q$、$1 - q = p$ であるから、

$$\mathrm{E}(X) = 1 + \frac{p}{q} + \frac{q}{p} = 1 + \frac{p^2 + q^2}{pq}.$$

さらに $p^2 + q^2 = (p+q)^2 - 2pq = 1 - 2pq$ であるから、

$$\mathrm{E}(X) = 1 + \frac{1 - 2pq}{pq} = 1 + \frac{1}{pq} - 2 = \frac{1}{pq} - 1,$$

となり、求める結果が得られる。

最後の小問

正の数 p と q に対する相加平均・相乗平均の不等式（AM–GM inequality）より、

$$\sqrt{pq} \leq \frac{p+q}{2} = \frac{1}{2}, \qquad \text{ゆえに} \qquad 0 < pq \leq \frac{1}{4}.$$

考えてみよう 逆数をとると、不等号の向きはどうなるだろうか。

$pq > 0$ であるから、逆数をとると不等号の向きが逆になり、$\dfrac{1}{pq} \geq 4$ である。したがって

$$\mathrm{E}(X) = \frac{1}{pq} - 1 \geq 4 - 1 = 3,$$

であり、等号が成り立つのは $p = q = \frac{1}{2}$ のときであり、かつそのときに限る。（同値な見方として、$pq = p(1-p) = \frac{1}{4} - (p - \frac{1}{2})^2 \leq \frac{1}{4}$ としてもよい。）

検算

- $p = q = \frac{1}{2}$ のとき：$X - 1$ は箱 1 と異なる中身の箱が初めて出るまでの待ち時間であり、平均 2 の幾何分布（geometric distribution）に従う確率変数である。よって $\mathrm{E}(X) = 3$ となり、$\frac{1}{1/4} - 1$ と一致する。

- $p \to 0$ のとき $\mathrm{E}(X) \to \infty$ となる。父ペンギンがめったに出ないので、長く待つことになる。これは $\frac{1}{pq} - 1$ が発散することと合っている。
- 直接計算すると、$n \geq 2$ に対して $\mathrm{P}(X = n) = p^{n-1}q + q^{n-1}p$ である。$p = \frac{1}{3}$ として $n\,\mathrm{P}(X = n)$ を数値的に足し合わせると 3.5 が得られる。また、$p = 0.3$ で 200 000 人の収集者をシミュレーションしたところ、平均は 3.761 となり、厳密値 $\frac{1}{0.21} - 1 \approx 3.762$ とよく一致した。

よくある間違い

- $p^{n-1} + q^{n-1}$ を $n = 1$ にもそのまま使ってしまうこと。これでは「確率」が 2 になり、期待値が 1 だけ大きくなってしまう。
- $\mathrm{P}(X \geq 4) = p^4 + q^4$ と書いてしまうこと。4 個目の箱が必要になるのは、最初の 3 個の箱の中身がそろっているときである。
- 最後の小問で、正の数どうしの不等式では逆数をとると不等号の向きが逆になることを忘れること。

STEP のその先へ

裾和公式には連続版の双子がある。非負の確率変数（random variable）に対して $\mathrm{E}(X) = \int_0^\infty \mathrm{P}(X > x)\,\mathrm{d}x$ が成り立ち、これもまったく同じ順序の入れ替えによって証明される（一般の場合にはトネリの定理（Tonelli's theorem）がこれを正当化する）。ペンギンの問題は**クーポン収集問題**（coupon collector's problem）の 2 種類の場合である。等確率の N 種類がある場合、必要な箱の個数の期待値は $N(1 + \frac{1}{2} + \cdots + \frac{1}{N})$ であり、およそ $N \ln N$ である。これらの考え方はどちらも、Grimmett と Welsh の *Probability: An Introduction*（確率論入門）や、Blitzstein と Hwang の *Introduction to Probability*（確率論入門）に載っている。

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。