解答例

# STEP III 2001、第 12 問

確率・統計 ・ 指示変数による期待値

> **STEP III 2001、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/01-s3-q12

## 鍵となる考え方

小問 (i) では、黒玉の個数の分布（distribution）を求めようとしてはいけない。個数を黒玉 1 個ごとの指示変数（indicator）に分解し、期待値の線形性（linearity of expectation）を用いるのである。袋が空になるまで非復元抽出（without replacement）を続けることは、すべての玉を無作為に並べることにほかならないので、各指示変数の確率は対称性（symmetry）の議論に帰着する。小問 (ii) では白玉が取り出されるまで袋の中身は変わらないので、各白玉までの待ち時間は幾何分布（geometric distribution）に従う。そして問題文の「これを用いて（hence）」の部分は、白玉が 1 個取り除かれると白玉が 1 個少ない同じ状況に戻る、ということに気づけば解ける。

## 解答

### 小問 (i)

袋が空になるまで取り出しを続けると考える。どの取り出しも無作為であり、玉は戻されないので、$b+w$ 個の玉が取り出される順序はそれらの一様ランダムな並び（uniformly random ordering）である。最後の白玉が取り出されるまでに取り出された黒玉とは、この並びにおいて最後の白玉より前にある黒玉にちょうど一致する。

黒玉に $1, 2, \ldots, b$ と番号をつけ、黒玉 $j$ が最後の白玉より前に取り出されるとき $I_j = 1$、そうでないとき $I_j = 0$ とする。最後の白玉が取り出されたときまでに取り出された黒玉の個数を $N$ とすると、

$$N = I_1 + I_2 + \cdots + I_b, \qquad \text{したがって} \qquad \mathrm{E}(N) = \sum_{j=1}^{b} \mathrm{P}(I_j = 1)$$

が期待値の線形性より成り立つ（これには $I_j$ どうしの独立性（independence）は必要ない）。

**考えてみよう** 黒玉 $j$ がすべての白玉より後に来る確率はいくらか。

黒玉 $j$ と $w$ 個の白玉からなる $w+1$ 個の玉だけに注目する。すべての玉の一様ランダムな並びにおいて、これら $w+1$ 個の相対的な順序もまた一様ランダムであるから、黒玉 $j$ はそれらの中の $w+1$ 個の位置のどれにも等しい確率で来る。黒玉 $j$ がこの組の中で最後になる確率は $\frac{1}{w+1}$ であり、これはちょうど事象 $I_j = 0$ である。よって

$$\mathrm{P}(I_j = 1) = 1 - \frac{1}{w+1} = \frac{w}{w+1},$$

であり、

$$\boxed{\mathrm{E}(N) = \frac{bw}{w+1}.}$$

**小問 (ii)**

今度は黒玉は袋に戻されるので、白玉が取り出されるまで袋の中には常に黒玉 $b$ 個と白玉 $w$ 個が入っている。毎回中身が同じであるから、各取り出しは以前の取り出しとは独立に、確率 $p = \frac{b}{b+w}$ で黒、確率 $q = \frac{w}{b+w}$ で白となる。

最初の白玉が出るまでに黒玉を取り出す回数を $X$ とする。$X = n$ とは黒が $n$ 回続いた後に白が出ることであるから、

$$\mathrm{P}(X = n) = p^n q, \qquad n = 0, 1, 2, \ldots$$

$b = 0$ ならば $X = 0$ であり結果は自明なので、$b \geq 1$ とする。このとき $0 < p < 1$ であり、$|x| < 1$ に対して $\sum_{n\geq 1} n x^{n-1} = \frac{1}{(1-x)^2}$ （幾何級数（geometric series）を微分して得られる）を用いると、

$$\mathrm{E}(X) = \sum_{n=0}^{\infty} n p^n q = pq \sum_{n=1}^{\infty} n p^{n-1} = \frac{pq}{(1-p)^2} = \frac{p}{q} = \frac{b/(b+w)}{w/(b+w)} = \frac{b}{w},$$

となる。ここで $1 - p = q$ を用いた。これが求める結果である。

最後の白玉については、黒玉を取り出す総回数を $X_w + X_{w-1} + \cdots + X_1$ と書く。ここで $X_k$ は袋の中にちょうど $k$ 個の白玉があるあいだに黒玉を取り出す回数である。白玉が $k$ 個残っているあいだ、袋には黒玉 $b$ 個と白玉 $k$ 個が入っているので、この段階は $w$ を $k$ に置き換えた「最初の白玉」の問題にほかならず、いま示した結果から $\mathrm{E}(X_k) = \frac{b}{k}$ を得る。期待値の線形性より、最後の白玉が取り出されるまでに黒玉を取り出す回数の期待値は

$$\boxed{\sum_{k=1}^{w} \frac{b}{k} = b\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{w}\right).}$$

## 検算

$w = 1$ のときは、どちらの小問も直接確かめやすい。(i) では、1 個しかない白玉より前に各黒玉が来る確率は $\frac{1}{2}$ なので $\frac{b}{2}$ となり、$\frac{bw}{w+1}$ と一致する。(ii) では段階は 1 つだけであり、和はちょうど $\frac{b}{1} = b$ である。(i) で $w \to \infty$ とすると、ほとんどすべての黒玉が最後の白玉より前に出てくるはずであり、実際 $\frac{bw}{w+1} \to b$ となる。

(i) については、$b \leq 4$、$w \leq 4$ のすべての場合についてあらゆる並びを列挙して確かめた。$b = 3$、$w = 4$ のときはちょうど $\frac{12}{5}$ となる。(ii) については、$b = 3$、$w = 4$ で $2 \times 10^5$ 回のシミュレーションを行ったところ、平均はおよそ 6.245 であり、$3(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4}) = \frac{25}{4} = 6.25$ とよく合う。

## よくある間違い

- (i) で、取り出される黒玉どうしが互いに独立であると仮定してしまう。期待値の線形性には独立性が不要であり、だからこそ指示変数がこれほど有効なのである。
- (ii) で、白玉を取り出す回も数に含める幾何分布を用いてしまう。すると $\frac{1}{q} = \frac{b+w}{w}$ となり、答えより 1 だけ大きくなる。
- 最後の部分で、すべての段階の分母に $b + w$ を用いてしまう。袋の中の白玉の個数は段階ごとに 1 個ずつ減っていく。

## STEP のその先へ

後半の問題は、姿を変えたクーポン収集問題（coupon collector's problem）である。各段階の幾何分布に従う待ち時間を足し合わせると調和和（harmonic sum）が現れ、これは $\ln w$ のように増大する。小問 (i) のように個数を指示変数の和として書く方法は、確率論で最も強力な手法の一つであり、ランダムな置換（random permutation）における不動点（fixed point）、記録（record）、巡回（cycle）の個数の期待値を計算するときに再び出会うことになる。Blitzstein と Hwang による *Introduction to Probability*（確率論入門）には、これら両方の考え方の例が数多く載っている。

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。