解答例

STEP I 1994、第 13 問

確率・統計 ・ 指示変数

> **STEP I 1994、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/94-s1-q13

鍵となる考え方

個数を指示変数（indicator variable）（それぞれ 1 または 0 をとる）の和として書くと、難しい期待値（expectation）の計算が易しいものに変わる。各部分が独立（independent）であってもなくても、期待値は線形だからである。ここでの取り出しは明らかに独立ではないが、対称性（symmetry）により各 X_i の期待値が等しいことがわかる。次に X の分布から直接 $\mathrm{E}(X)$ を別の方法で計算し、二つの答えを等しいとおけば、二項係数（binomial coefficient）の恒等式がほとんど労せずに得られる。

解答

以下、つねに $1 \le n \le M$ とし、$b > a$ または $b < 0$ のとき $\binom{a}{b} = 0$ と約束する。

小問 (i)

各 X_i は、i 番目に取り出した玉が赤ならば 1、そうでなければ 0 である。したがって和 $X_1 + X_2 + \cdots + X_n$ において、取り出した赤い玉はそれぞれちょうど 1 だけ寄与し、赤でない玉は 0 だけ寄与する。よってこの和は取り出した n 個のうちの赤い玉の個数を数えており、それが X である。

小問 (ii)

X_i は 0 と 1 の値しかとらないから、

$$\mathrm{E}(X_i) = 0 \cdot \mathrm{P}(X_i = 0) + 1 \cdot \mathrm{P}(X_i = 1) = \mathrm{P}(i \text{ 番目に取り出す玉が赤}).$$

考えてみよう i 番目の玉が赤である確率は、それより前に何を取り出したかに依存するのではないか？

条件付きで考えれば依存するが、条件を付けなければ依存しない。玉に番号を付け、すべてを一つずつ取り出していくことを想像しよう。M 個の玉の並び順はどれも同様に確からしい。したがって i 番目の位置にある玉は、1 番目の位置の玉と同じく、M 個の玉のどれである確率も等しい。n 回で取り出しをやめても、位置 $i \le n$ で起こることは変わらない。M 個のうち m 個が赤であるから、

$$\mathrm{E}(X_i) = \mathrm{P}(X_i = 1) = \frac{m}{M} \qquad \text{各} i = 1, \ldots, n.$$

（$i = 2$ の場合は地道な計算でも確かめられる：$\frac{m}{M} \cdot \frac{m-1}{M-1} + \frac{M-m}{M} \cdot \frac{m}{M-1} = \frac{m(M-1)}{M(M-1)} = \frac{m}{M}$。）

小問 (iii)

小問 (i) と期待値の線形性（linearity of expectation）（独立性は必要ない）により、

$$\mathrm{E}(X) = \mathrm{E}(X_1) + \cdots + \mathrm{E}(X_n) = n \cdot \frac{m}{M} = \frac{mn}{M},$$

となり、示すべきことが得られた。

小問 (iv)

取り出される n 個の玉の集合は、大きさ n の $\binom{M}{n}$ 個の部分集合のどれである確率も等しい。赤い玉がちょうど k 個であるためには、m 個の赤い玉から k 個、残りの $M - m$ 個から $n - k$ 個を選べばよいので、

$$\mathrm{P}(X = k) = \frac{\binom{m}{k}\binom{M-m}{n-k}}{\binom{M}{n}}, \qquad k = 0, 1, \ldots, n.$$

二項係数についての約束により、$k > m$（赤い玉が足りない）や $n - k > M - m$（それ以外の玉が足りない）のときには、この値は正しく 0 となる。これが**超幾何分布**（hypergeometric distribution）である。

小問 (v)

小問 (iv) から $\mathrm{E}(X)$ を直接計算する。$k = 0$ の項は 0 であるから、

$$\mathrm{E}(X) = \sum_{k=1}^{n} k\,\mathrm{P}(X = k) = \frac{1}{\binom{M}{n}} \sum_{k=1}^{n} k \binom{m}{k}\binom{M-m}{n-k}.$$

これを小問 (iii) の結果と等しいとおくと、

$$\sum_{k=1}^{n} k \binom{m}{k}\binom{M-m}{n-k} = \frac{mn}{M}\binom{M}{n}.$$

最後に、

$$\frac{n}{M}\binom{M}{n} = \frac{n}{M}\cdot\frac{M!}{n!\,(M-n)!} = \frac{(M-1)!}{(n-1)!\,(M-n)!} = \binom{M-1}{n-1},$$

であるから、右辺は $m\binom{M-1}{n-1}$ となり、これが求める恒等式である。

検算

- $M=5$、$m=2$、$n=3$ とする。小問 (iv) より $\mathrm{P}(X=0,1,2) = \frac{1}{10}, \frac{6}{10}, \frac{3}{10}$ であり、その和は 1 である。また $\mathrm{E}(X) = \frac{6+6}{10} = \frac{6}{5} = \frac{mn}{M}$ となる。恒等式は $1\cdot2\cdot3+2\cdot1\cdot3 = 12 = 2\binom{4}{2}$ となる。
- $n=M$（すべて取り出す）ならば確実に $X=m$ であり、実際 $\mathrm{E}(X) = mM/M = m$ となる。
- $M \le 6$ のすべての袋について玉の並び順をすべて列挙して $\mathrm{E}(X_i)$、$\mathrm{E}(X)$、$\mathrm{P}(X=k)$ を確かめ、また $M<30$ のすべてについて (v) の恒等式を確かめた。

よくある間違い

- $\mathrm{E}(X) = \sum \mathrm{E}(X_i)$ には独立性が必要だと主張すること。必要ない。期待値の線形性はつねに成り立つ。
- 取り出しを独立とみなし、(iv) で二項分布 $\mathrm{B}(n, \frac{m}{M})$ と書くこと。これでは平均は正しくなるが確率は誤りとなる。取り出しは非復元抽出（without replacement）だからである。

STEP のその先へ

指示変数の方法は確率論で最も役に立つ道具の一つである。ランダムな置換（permutation）の固定点（fixed point）の個数、硬貨投げにおける連（run）の個数、ランダムグラフ（random graph）の孤立点（isolated vertex）の個数などの期待値を与え、さらに $\mathrm{Var}(\sum X_i) = \sum \mathrm{Var}(X_i) + \sum_{i\neq j} \mathrm{Cov}(X_i, X_j)$ によって分散（variance）にも拡張できる。一つの量を二通りに計算して組合せの恒等式を証明する方法も、同じく広く使える。Blitzstein と Hwang の *Introduction to Probability*（確率論入門）は全編を通して指示変数を用いており、この二つの考え方が実際に働く様子を見るのに格好の本である。

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。