解答例

# STEP I 2018、第 12 問

確率・統計 ・ 復元抽出

> **STEP I 2018、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/18-s1-q12

## 鍵となる考え方

表の出た回数を、1 回ごとの投げに対応する 2 つの指示変数（indicator variable）の和として書く。和の期待値（expectation）は、各回の投げが独立（independent）であるかどうかにかかわらず期待値の和に等しいので、平均はたやすく求まる。分散（variance）はただ 1 つの数、すなわち 2 回の投げがともに表となる確率で決まり、復元抽出（drawing with replacement）と非復元抽出（drawing without replacement）が違ってくるのはまさにこの点である。

## 解答

以下を通して、$p = \frac{1}{3}(p_1 + p_2 + p_3)$ および

$$q = p_1p_2 + p_2p_3 + p_3p_1.$$

とおく。どの硬貨も等しい確率で選ばれ、硬貨が選ばれたあとの投げは互いに独立であると仮定する。

### 小問 (i)

硬貨 $k$ が選ばれる事象を $C_k$ とすると、$\mathrm{P}(C_k) = \frac{1}{3}$ である。全確率の法則（law of total probability）より、

$$\mathrm{P}(\text{表}) = \sum_{k=1}^{3} \mathrm{P}(C_k)\,\mathrm{P}(\text{表} \mid C_k) = \tfrac{1}{3}(p_1 + p_2 + p_3) = p.$$

**小問 (ii)**

1 回目と 2 回目の投げで表が出ることの指示変数をそれぞれ $X_1$、$X_2$ とすると、$N_1 = X_1 + X_2$ である。1 枚目の硬貨は戻されるので、2 回目の投げは 1 回目とまったく同じ状況で行われ、小問 (i) より $\mathrm{P}(X_1 = 1) = \mathrm{P}(X_2 = 1) = p$ である。したがって

$$\mathrm{E}(N_1) = \mathrm{E}(X_1) + \mathrm{E}(X_2) = 2p.$$

分散を求めるには $\mathrm{P}(X_1 = 1,\, X_2 = 1)$ が必要である。2 回の選択は独立であり、硬貨の順序対 $(j,k)$ 9 通りはそれぞれ確率 $\frac{1}{9}$ をもち、その対が与えられたとき 2 回とも表となる確率は $p_j p_k$ である。よって

$$\mathrm{P}(X_1 = 1,\, X_2 = 1) = \frac{1}{9}\sum_{j=1}^{3}\sum_{k=1}^{3} p_j p_k = \left(\frac{p_1 + p_2 + p_3}{3}\right)^2 = p^2 = \mathrm{P}(X_1 = 1)\,\mathrm{P}(X_2 = 1).$$

他の結果についても同じように積に分解できるので、$X_1$ と $X_2$ は独立であり、$N_1 \sim \mathrm{B}(2, p)$ となる。直接計算することもできる。$X_i^2 = X_i$ であるから、

$$\begin{aligned}\mathrm{E}(N_1^2) &= \mathrm{E}(X_1) + \mathrm{E}(X_2) + 2\,\mathrm{E}(X_1 X_2) = 2p + 2p^2,\\ \mathrm{Var}(N_1) &= 2p + 2p^2 - (2p)^2 = 2p(1-p),\end{aligned}$$

となり、求める結果が得られる。

**小問 (iii)**

今度は 2 枚の硬貨を戻さずに 1 枚ずつ順に選び、1 枚目と 2 枚目に選んだ硬貨で表が出ることの指示変数をそれぞれ $Y_1$、$Y_2$ とすると、$N_2 = Y_1 + Y_2$ である。どの硬貨も 1 枚目に選ばれる確率は等しく、対称性からどの硬貨も 2 枚目に選ばれる確率は等しいので、小問 (i) と同様に $\mathrm{P}(Y_1 = 1) = \mathrm{P}(Y_2 = 1) = p$ であり、

$$\mathrm{E}(N_2) = 2p.$$

? **考えてみよう** 小問 (ii) での P(2 回とも表) の計算をそのまま使えるだろうか？

使えない。同じ硬貨を 2 回選ぶことはもはやできないからである。今度は $j \neq k$ を満たす順序対 $(j,k)$ が 6 通りあって同様に確からしく、各非順序対 $\{j,k\}$ は 2 回ずつ現れるので、

$$\mathrm{P}(Y_1 = 1,\, Y_2 = 1) = \frac{1}{6}\sum_{j\neq k} p_j p_k = \frac{2q}{6} = \frac{q}{3}.$$

これは一般には $p^2$ に等しくないので、$Y_1$ と $Y_2$ は独立ではない。再び $Y_i^2 = Y_i$ を用い

て、

$$\mathrm{E}(N_2^2) = \mathrm{E}(Y_1) + \mathrm{E}(Y_2) + 2\,\mathrm{E}(Y_1Y_2) = 2p + \tfrac{2}{3}q,$$
$$\mathrm{Var}(N_2) = 2p + \tfrac{2}{3}(p_1p_2 + p_2p_3 + p_3p_1) - 4p^2.$$

**小問 (iv)**

差をとると、

$$\mathrm{Var}(N_1) - \mathrm{Var}(N_2) = (2p - 2p^2) - (2p + \tfrac{2}{3}q - 4p^2) = 2p^2 - \tfrac{2}{3}q.$$

ここで $9p^2 = (p_1 + p_2 + p_3)^2 = p_1^2 + p_2^2 + p_3^2 + 2q$ であるから、

$$\begin{aligned} 2p^2 - \tfrac{2}{3}q &= \tfrac{2}{9}(p_1^2 + p_2^2 + p_3^2 + 2q) - \tfrac{6}{9}q = \tfrac{2}{9}(p_1^2 + p_2^2 + p_3^2 - p_1p_2 - p_2p_3 - p_3p_1) \\ &= \tfrac{1}{9}\Big[(p_1 - p_2)^2 + (p_2 - p_3)^2 + (p_3 - p_1)^2\Big], \end{aligned}$$

となる。最後の等号は、3 つの平方を展開すると $2(p_1^2 + p_2^2 + p_3^2) - 2q$ となることによる。

実数の平方の和は負にならないので、$\mathrm{Var}(N_2) \le \mathrm{Var}(N_1)$ である。等号が成り立つのは 3 つの平方がすべて 0 となるとき、すなわち $p_1 = p_2 = p_3$ のとき、かつそのときに限る。以上で証明が完了する。

副産物として、小問 (iii) の答えのすっきりした形が得られる：

$$\mathrm{Var}(N_2) = 2p(1-p) - \tfrac{1}{9}\Big[(p_1 - p_2)^2 + (p_2 - p_3)^2 + (p_3 - p_1)^2\Big].$$

## 検算

- $p_1 = p_2 = p_3 = p$ ならば、どの硬貨も同じなので、硬貨の選び方は関係ないはずである。実際、どちらの回数も $\mathrm{B}(2, p)$ に従い、分散は一致して、等号成立の場合と合っている。
- $p_1 = \frac{1}{2}$、$p_2 = \frac{1}{3}$、$p_3 = \frac{1}{6}$ とすると、$p = \frac{1}{3}$、$q = \frac{11}{36}$ である。このとき $\mathrm{Var}(N_1) = \frac{4}{9}$、$\mathrm{Var}(N_2) = \frac{2}{3} + \frac{11}{54} - \frac{4}{9} = \frac{23}{54}$ であり、差は $\frac{1}{54}$ で、$\frac{1}{9}\left(\frac{1}{36} + \frac{1}{36} + \frac{1}{9}\right) = \frac{1}{54}$ と一致する。すべての結果を手で書き出しても同じ値が得られる。
- 極端な場合 $p_1 = p_2 = 1$、$p_3 = 0$：小問 (iii) では、硬貨 1 と 2 がともに選ばれる場合を除いて、選ばれた 2 枚のうちちょうど 1 枚が必ず表を出す硬貨である。よって $N_2$ は確率 $\frac{1}{3}$ で 2、それ以外では 1 となり、分散は $\frac{2}{9}$ である。公式からも $2p + \frac{2}{3}q - 4p^2 = \frac{4}{3} + \frac{2}{3} - \frac{16}{9} = \frac{2}{9}$ が得られる。

## よくある間違い

- 小問 (iii) の 2 回の投げを独立とみなして、再び $2p(1-p)$ を得てしまう。小問 (iv) の要点はまさに、それらが独立でないことにある。
- $\mathrm{Var}(N_1)$ を各硬貨の分散 $2p_k(1-p_k)$ の平均として書いてしまう。これは硬貨を無作為に選ぶことから生じる余分なばらつきを無視しており、しかも小問 (ii) では投げるたびに異なる硬貨が使われうる。

## STEP のその先へ

小問 (ii) と (iii) の違いは、復元抽出と非復元抽出の違いである。非復元抽出では各回の抽出が負の相関（negative correlation）をもち、それによって分散が小さくなる。同じ効果が標本調査における「有限母集団修正（finite population correction）」を生み、また超幾何分布（hypergeometric distribution）が同じ平均をもつ二項分布（binomial distribution）よりも分散が小さい理由にもなっている。小問 (iv) の代数計算は恒等式 $\sum p_k^2 - \sum_{j<k} p_j p_k = \frac{1}{2}\sum_{j<k}(p_j - p_k)^2$ であり、分散が決して負にならないという事実の近い親戚である。Blitzstein と Hwang の *Introduction to Probability*（確率論入門）は、指示変数、共分散（covariance）、超幾何分布をまさにこの精神で扱っている。

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。