詳細解答

STEP I 2018，第 12 題

概率與統計 • 放回抽樣

> **STEP I 2018，第 12 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/18-s1-q12

關鍵思路

把每個正面次數寫成兩個指示變量（indicator variable）之和，每次拋擲對應一個。無論各次拋擲是否獨立（independent），和的期望值（expectation）都等於期望值之和，所以平均值很容易求得。方差（variance）則取決於一個數，就是兩次拋擲皆出現正面的概率，而這正是放回抽樣（drawing with replacement）與不放回抽樣（drawing without replacement）的分別所在。

解答

全文中，設 $p = \frac{1}{3}(p_1 + p_2 + p_3)$ 及

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

我們假設每枚硬幣被選中的機會相等，而且硬幣選定後，各次拋擲互相獨立。

第 (i) 部分

設 C_k 為選中第 k 枚硬幣的事件，故 $\mathrm{P}(C_k) = \frac{1}{3}$。由全概率定律（law of total probability），

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

第 (ii) 部分

設 X_1 和 X_2 分別為第一次和第二次拋擲出現正面的指示變量，故 $N_1 = X_1 + X_2$。由於第一枚硬幣會被放回，第二次拋擲的情況與第一次完全相同，由第 (i) 部分得 $\mathrm{P}(X_1 =$

$1) = \mathrm{P}(X_2 = 1) = p$。因此

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

求方差需要 $\mathrm{P}(X_1 = 1,\, X_2 = 1)$。兩次選擇互相獨立，9 個有序硬幣對 (j,k) 的概率各為 $\frac{1}{9}$，而給定該硬幣對時，兩次皆為正面的機會是 $p_j p_k$。所以

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

同樣的因式分解對其他結果亦成立，因此 X_1 與 X_2 互相獨立，且 $N_1 \sim \mathrm{B}(2, p)$。也可以直接計算：由於 $X_i^2 = X_i$，

$$\begin{aligned}\mathrm{E}(N_1^2) &= \mathrm{E}(X_1) + \mathrm{E}(X_2) + 2\,\mathrm{E}(X_1X_2) = 2p + 2p^2,\\ \mathrm{Var}(N_1) &= 2p + 2p^2 - (2p)^2 = 2p(1-p),\end{aligned}$$

正如所求。

第 (iii) 部分

現在以不放回方式先後選出兩枚硬幣，設 Y_1、Y_2 分別為第一枚和第二枚選出的硬幣出現正面的指示變量，故 $N_2 = Y_1 + Y_2$。每枚硬幣成為第一枚被選中的機會相等，由對稱性，每枚成為第二枚被選中的機會亦相等，所以與第 (i) 部分一樣，$\mathrm{P}(Y_1 = 1) = \mathrm{P}(Y_2 = 1) = p$，且

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

> **想一想** 可以沿用第 (ii) 部分中 P(兩次皆正面) 的計算嗎？

不可以：同一枚硬幣不能再被選中兩次。現在有 6 個機會相等的有序對 (j,k)，其中 $j \neq k$，而每個無序對 $\{j,k\}$ 出現兩次，所以

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

一般而言這並不等於 p^2，故 Y_1 與 Y_2 並不獨立。再次利用 $Y_i^2 = Y_i$，

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

第 (iv) 部分

兩式相減，

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

而 $9p^2 = (p_1+p_2+p_3)^2 = p_1^2+p_2^2+p_3^2+2q$，所以

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

最後一步成立，是因為展開這三個平方會得到 $2(p_1^2+p_2^2+p_3^2)-2q$。

實數的平方和是非負的，所以 $\mathrm{Var}(N_2) \le \mathrm{Var}(N_1)$。等號成立當且僅當三個平方皆為零，即當且僅當 $p_1 = p_2 = p_3$。證明完畢。

作為副產品，我們得到第 (iii) 部分答案的一個簡潔形式：

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

驗算

- 若 $p_1 = p_2 = p_3 = p$，每枚硬幣都一樣，那麼硬幣怎樣選出並無分別：兩個次數都服從 $\mathrm{B}(2,p)$，方差相等，與等號成立的情況吻合。
- 取 $p_1 = \frac{1}{2}$、$p_2 = \frac{1}{3}$、$p_3 = \frac{1}{6}$，則 $p = \frac{1}{3}$，$q = \frac{11}{36}$。於是 $\mathrm{Var}(N_1) = \frac{4}{9}$，$\mathrm{Var}(N_2) = \frac{2}{3} + \frac{11}{54} - \frac{4}{9} = \frac{23}{54}$，少了 $\frac{1}{54}$，與 $\frac{1}{9}\left(\frac{1}{36}+\frac{1}{36}+\frac{1}{9}\right) = \frac{1}{54}$ 一致。親手列出所有結果亦得出相同數值。
- 極端情況 $p_1 = p_2 = 1$、$p_3 = 0$：在第 (iii) 部分中，除非第 1 和第 2 枚硬幣同時被選中，否則選出的兩枚硬幣中恰有一枚必定是正面，因此 N_2 以概率 $\frac{1}{3}$ 取值 2，其餘情況取值 1，方差為 $\frac{2}{9}$。公式給出 $2p + \frac{2}{3}q - 4p^2 = \frac{4}{3} + \frac{2}{3} - \frac{16}{9} = \frac{2}{9}$。

常見錯誤

- 把第 (iii) 部分的兩次拋擲當作獨立，於是又得到 $2p(1-p)$；第 (iv) 部分的重點正正在於它們並不獨立。
- 把 $\mathrm{Var}(N_1)$ 寫成各枚硬幣方差 $2p_k(1-p_k)$ 的平均；這忽略了隨機選擇硬幣所帶來的額外變異，而且在第 (ii) 部分中，每次拋擲可能用不同的硬幣。

STEP 之外

第 (ii) 與第 (iii) 部分的分別，就是放回抽樣與不放回抽樣的分別。不放回抽樣使各次抽取呈負相關（negatively correlated），從而降低方差；同樣的效應在抽樣調查中產生「有限總體校正」（finite population correction），亦解釋了為何超幾何分佈（hypergeometric distribution）的方差比具相同平均值的二項分佈（binomial distribution）小。第 (iv) 部

分的代數運算就是恆等式 $\sum p_k^2 - \sum_{j<k} p_j p_k = \frac{1}{2}\sum_{j<k}(p_j - p_k)^2$，它與「方差永不為負」這個事實關係密切。Blitzstein 和 Hwang 的 *Introduction to Probability*（《概率導論》）正是以這種精神講解指示變量、協方差（covariance）和超幾何分佈。

Beyond STEP 詳細解答：https://beyondstep.keithmcnulty.org/zh-hant/。解答文字採用 CC0 1.0；題目本身版權屬 Cambridge University Press & Assessment 所有（©），本文件不予轉載。由機器從英文翻譯而成；尚未經審閱。