詳細解答

STEP I 2003，第 12 題

概率與統計 • 乘積的期望值

> **STEP I 2003，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/03-s1-q12

關鍵思路

期望值 (expected value) 就是對所有等可能的結果，把 (數值 × 概率) 加起來，所以兩部分都歸結為列出所有有序數對（ordered pairs），再把它們的乘積相加。二重和 $\sum_r \sum_s rs$ 可以分解為 $(\sum r)^2$，第 (i) 部分要做的就只有這一步。第 (ii) 部分唯一的分別，是 $r = s$ 的數對不再可能出現，所以我們取同一個二重和，再除去它的對角線項。

解答

設 X 為第一個球上的數字，Y 為第二個球上的數字。

第 (i) 部分

由於第一個球會放回袋中，第二次抽取時袋中仍是全部的球，結果與第一次無關。所以對任意 $r, s \in \{1, \ldots, n\}$，

$$\mathrm{P}(X = r,\ Y = s) = \mathrm{P}(X = r)\,\mathrm{P}(Y = s) = \frac{1}{n} \cdot \frac{1}{n} = \frac{1}{n^2}.$$

這樣的有序數對 (r, s) 共有 n^2 個，每個給出乘積 rs，所以按期望值的定義，

$$\mathrm{E}(XY) = \sum_{r=1}^{n} \sum_{s=1}^{n} rs\,\mathrm{P}(X = r,\ Y = s) = \frac{1}{n^2} \sum_{r=1}^{n} \sum_{s=1}^{n} rs,$$

這正是題目要求我們解釋的式子。

要化簡這個式子，留意 r 與 s 無關，所以可以把它提到內層求和之外：

$$\sum_{r=1}^{n} \sum_{s=1}^{n} rs = \sum_{r=1}^{n} r\Big(\sum_{s=1}^{n} s\Big) = \Big(\sum_{r=1}^{n} r\Big)^2 = \tfrac{1}{4} n^2 (n+1)^2.$$

因此

$$\mathrm{E}(XY) = \frac{(n+1)^2}{4}.$$

（等價地說，X 與 Y 獨立（independent），所以 $\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y) = (\frac{n+1}{2})^2$。）

第 (ii) 部分

要抽出兩個球，須有 $n \geq 2$。現在第一個球不放回，所以可能的結果是滿足 $r \neq s$ 的有序數對 (r,s)。對這樣的一個數對，

$$\mathrm{P}(X = r,\ Y = s) = \mathrm{P}(X = r)\,\mathrm{P}(Y = s \mid X = r) = \frac{1}{n} \cdot \frac{1}{n-1},$$

因為取走 r 號球之後，第二個球是餘下 $n-1$ 個球中任何一個的可能性均等。這樣的數對共有 $n(n-1)$ 個，全部等可能，而 $r = s$ 的數對概率為 0。因此

$$\mathrm{E}(XY) = \frac{1}{n(n-1)} \sum_{r \neq s} rs.$$

想一想 怎樣只對 $r \neq s$ 的數對把 rs 加起來？

先取對所有數對的和（已在第 (i) 部分求出），再減去 $r = s$ 的對角線項，這些項各自貢獻 $r \cdot r = r^2$：

$$\sum_{r \neq s} rs = \sum_{r=1}^{n} \sum_{s=1}^{n} rs - \sum_{r=1}^{n} r^2 = \tfrac{1}{4}n^2(n+1)^2 - \tfrac{1}{6}n(n+1)(2n+1).$$

提出公因式 $\frac{1}{12}n(n+1)$，

$$\begin{aligned} \sum_{r \neq s} rs &= \tfrac{1}{12}n(n+1)[3n(n+1) - 2(2n+1)] \\ &= \tfrac{1}{12}n(n+1)(3n^2 - n - 2) \\ &= \tfrac{1}{12}n(n+1)(n-1)(3n+2). \end{aligned}$$

除以 $n(n-1)$（因為 $n \geq 2$，它不等於零），

$$\mathrm{E}(XY) = \frac{(n+1)(3n+2)}{12},$$

這就是所求的結果。

驗算

取 $n = 3$。有放回時，九個乘積的平均值是 $\frac{1}{9}(1+2+3)^2 = 4$，而 $\frac{(3+1)^2}{4} = 4$。無放回時，無序數對 $\{1,2\}, \{1,3\}, \{2,3\}$ 給出乘積 $2, 3, 6$，平均值為 $\frac{11}{3}$，而 $\frac{4 \cdot 11}{12} = \frac{11}{3}$。當 $n = 2$

時，無放回的乘積只有 2 一個，而公式給出 $\frac{3\cdot 8}{12} = 2$。我們亦以逐一列舉的方式，對 12 或以下的每個 n 準確驗證了兩條公式。

比較兩個答案也是一個很好的合理性檢查：

$$\frac{(n+1)(3n+2)}{12} - \frac{(n+1)^2}{4} = -\frac{n+1}{12} < 0.$$

無放回時答案較小，這是合理的：我們失去了 $n \cdot n$ 這類數值大的對角線乘積。

常見錯誤

- 在第 (ii) 部分除以 n^2。可能的有序數對只有 $n(n-1)$ 個，每個的概率是 $\frac{1}{n(n-1)}$。
- 求和時用無序數對，計數時卻用有序數對（或者反過來）。對無序數對求和，會令總和與個數同時減半，所以比值不變；但把兩者混用，結果就會相差一個因子 2。
- 在第 (ii) 部分寫 $\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y)$。這需要獨立性，而球不放回時 X 與 Y 並不獨立。

STEP 之外

我們求得的差正好就是協方差（covariance）：$\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y) = -\frac{n+1}{12}$，它是負數，因為第一個數字大，第二個數字也大的可能性就會稍低。在無放回抽樣（sampling without replacement）中求和的方差（variance）時，需要的正是這類協方差，由此可以得出超幾何分佈（hypergeometric distribution）的方差，以及抽樣調查中使用的「有限總體校正」（finite population correction）。Blitzstein 與 Hwang 的 *Introduction to Probability*（概率論入門教材）正是以這種方式發展協方差與指示變量（indicator）方法。

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。