詳細解答

STEP III 2021，第 12 題

概率與統計 • 和的方差

> **STEP III 2021，第 12 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 OCR 提供下載的 2021 年 STEP 試卷（zip 檔案）內的 STEP III 試卷：
>
> www.ocr.org.uk/Images/696588-2021-past-papers.zip

本題在 *Beyond STEP* 第 12.4 節兩兩獨立，卻非相互獨立中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch12/pairwise-independence.html

關鍵思路

把全隊的得分寫成一些指示型變量（indicator variable）之和，每對球員對應一個變量，然後展開這個和的方差（variance）。一切都取決於兩個不同的球員對之間如何相互影響。如果兩對沒有共同的球員，它們顯然是獨立（independent）的，所以真正要處理的是共有一名球員的兩對。在 (i) 中，這樣的兩對原來是獨立的，所以交叉項消失。在 (iii) 中則不然，而 (ii) 中的恆等式正是計算它們貢獻的工具。

解答

第 (i) 部分

每顆骰子都是公正的，各球員獨立擲骰。不論兩名指定球員中的第一人擲出甚麼，二人擲出相同點數的概率都是 $\frac{1}{6}$。所以

$$\mathrm{P}(X_{12}=1)=\mathrm{P}(X_{23}=1)=\tfrac{1}{6}.$$

$X_{12}=1$ 和 $X_{23}=1$ 同時發生，當且僅當球員 1、2、3 都擲出相同點數，所以

$$\mathrm{P}(X_{12}=1,\ X_{23}=1)=\tfrac{1}{6}\cdot\tfrac{1}{6}=\tfrac{1}{36}=\mathrm{P}(X_{12}=1)\,\mathrm{P}(X_{23}=1).$$

從邊際概率（marginal probability）中減去這個值，便得到其餘三格。例如 $\mathrm{P}(X_{12}=1,X_{23}=0)=\frac{1}{6}-\frac{1}{36}=\frac{1}{6}\cdot\frac{5}{6}$；同理 $\mathrm{P}(X_{12}=0,X_{23}=1)=\frac{5}{6}\cdot\frac{1}{6}$，以及 $\mathrm{P}(X_{12}=0,X_{23}=$

$0) = 1 - \frac{1}{6} - \frac{1}{6} + \frac{1}{36} = \frac{5}{6} \cdot \frac{5}{6}$。所以聯合分佈（joint distribution）等於邊際分佈的乘積，X_{12} 與 X_{23} 是獨立的。

> **想一想** 為甚麼知道 $X_{12} = 1$ 對 X_{23} 毫無資訊？

已知球員 1 與 2 點數相同，球員 2 的點數在 $\{1, \ldots, 6\}$ 上仍然是均勻分佈（uniform）的，而球員 3 與之相同的概率是 $\frac{1}{6}$，跟之前完全一樣。由於每一面出現的機會均等，共有的那名球員並不帶來任何資訊。

設 $T = \sum_{i<j} X_{ij}$ 為總得分，它是 $N = \binom{n}{2}$ 項之和。每個 X_{ij} 都是參數為 $\frac{1}{6}$ 的伯努利（Bernoulli）變量，所以 $\mathrm{E}(X_{ij}) = \frac{1}{6}$，$\mathrm{Var}(X_{ij}) = \frac{1}{6} \cdot \frac{5}{6} = \frac{5}{36}$。因此

$$\mathrm{E}(T) = \frac{N}{6} = \frac{n(n-1)}{12}.$$

至於方差，任取兩個不同的球員對。如果它們沒有共同球員，它們取決於互不相交的兩組骰子，所以是獨立的。如果它們共有一名球員，那麼把球員重新編號後，由上面的論證可知它們也是獨立的。所以各 X_{ij} 是成對獨立（pairwise independent）的，每個協方差（covariance）都是零，和的方差等於方差之和：

$$\mathrm{Var}(T) = N \cdot \frac{5}{36} = \frac{5n(n-1)}{72}.$$

第 (ii) 部分

設 $S = Y_1 + \cdots + Y_m$。由線性（linearity）得 $\mathrm{E}(S) = 0$，所以 $\mathrm{Var}(S) = \mathrm{E}(S^2) - 0^2 = \mathrm{E}(S^2)$。展開平方，每個 Y_i^2 出現一次，而每個 $i < j$ 的乘積 $Y_i Y_j$ 出現兩次：

$$S^2 = \sum_{i=1}^{m} Y_i^2 + 2 \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} Y_i Y_j.$$

取期望值（expectation）並利用線性，便得到所求的公式。

第 (iii) 部分

對兩名球員而言，$\mathrm{P}(\text{擲出相同的偶數}) = \frac{3}{36} = \frac{1}{12}$，同樣 $\mathrm{P}(\text{擲出相同的奇數}) = \frac{1}{12}$。所以

$$\mathrm{P}(Z_{ij} = 1) = \mathrm{P}(Z_{ij} = -1) = \tfrac{1}{12}, \qquad \mathrm{P}(Z_{ij} = 0) = \tfrac{5}{6}.$$

不獨立。如果 $Z_{12} = 1$ 且 $Z_{23} = -1$，球員 2 的點數便要既是偶數又是奇數。所以

$$\mathrm{P}(Z_{12} = 1,\ Z_{23} = -1) = 0 \neq \tfrac{1}{12} \cdot \tfrac{1}{12} = \mathrm{P}(Z_{12} = 1)\,\mathrm{P}(Z_{23} = -1),$$

因此 Z_{12} 與 Z_{23} 並不獨立。

平均值。由對稱性，$\mathrm{E}(Z_{ij}) = \frac{1}{12} - \frac{1}{12} = 0$，所以總得分 $U = \sum_{i<j} Z_{ij}$ 的平均值為 0。

方差。每個 Z_{ij} 的平均值都是零，所以可以對這 $N = \binom{n}{2}$ 個變量 Z_{ij} 應用 (ii)。我們需要兩類期望值。

- 平方項：Z_{ij}^2 等於 1 當且僅當球員 i 與 j 點數相同，所以 $\mathrm{E}(Z_{ij}^2) = \frac{1}{6}$。
- 兩個不同球員對的乘積。如果兩對沒有共同球員，兩個變量獨立且平均值為零，所以 $\mathrm{E}(Z_{ij}Z_{kl}) = \mathrm{E}(Z_{ij})\,\mathrm{E}(Z_{kl}) = 0$。如果它們共有一名球員，例如 $Z_{12}Z_{23}$，那麼只有當三名球員都擲出相同點數時乘積才非零。這時兩個因子同為 $+1$（偶數）或同為 -1（奇數），所以乘積是 1。因此 $\mathrm{E}(Z_{12}Z_{23}) = \frac{1}{36}$。

想一想 有多少個無序的「球員對之對」恰好共有一名球員？

先以 n 種方法選出共有的球員，再從其餘 $n-1$ 名球員中以 $\binom{n-1}{2}$ 種方法選出他的兩個搭檔。這樣得到 $n\binom{n-1}{2} = \frac{1}{2}n(n-1)(n-2)$ 個這樣的「球員對之對」，每個只計算一次，正是 (ii) 中雙重求和所需要的。

綜合起來，

$$\begin{aligned}\mathrm{Var}(U) &= \binom{n}{2}\cdot\frac{1}{6} + 2\cdot\frac{n(n-1)(n-2)}{2}\cdot\frac{1}{36}\\ &= \frac{n(n-1)}{36}\big(3+(n-2)\big)\\ &= \frac{n(n-1)(n+1)}{36} = \frac{1}{36}n(n^2-1),\end{aligned}$$

得證。

驗算

- 以例子中的擲骰結果 $3,3,3,3,6,6,2$ 為例，配對計數得 $6+1=7$，帶符號計數得 $-6+1=-5$，與題目一致。
- 當 $n=2$ 時只有一對，所以 $\mathrm{Var}(U) = \mathrm{E}(Z_{12}^2) = \frac{1}{6}$，而公式給出 $\frac{2\cdot3}{36} = \frac{1}{6}$。
- 當 $n=3$ 時，若三顆骰子點數全部相同，則 $U = \pm3$（各自概率為 $\frac{1}{72}$）。否則恰好有一對點數相同，$U = \pm1$，總概率為 $\frac{15}{36}$。所以 $\mathrm{E}(U^2) = 9\cdot\frac{1}{36} + \frac{15}{36} = \frac{2}{3}$，與 $\frac{3\cdot8}{36}$ 相符。以電腦精確列舉 $n \le 6$ 的所有結果，證實了兩個平均值和兩個方差；例如 $n=6$ 時，(i) 得 $\frac{25}{12}$，(iii) 得 $\frac{35}{6}$。

常見錯誤

- 只檢查了一格便聲稱 X_{12} 與 X_{23} 獨立，卻沒有說明為何其他格也隨之成立。對於只取兩個值的變量，其他格確實隨之成立，但要把這一點說出來。

- 在 (iii) 中，忘記了共有一名球員的兩對不論共同點數是偶數還是奇數，貢獻都是 $+\frac{1}{36}$。負號在乘積中互相抵消。
- 數錯有重疊的「球員對之對」的數目，例如用了 $n(n-1)(n-2)$（有序）之後再乘以 2。

STEP 之外

第 (i) 部分是一個經典例子：變量成對獨立，卻並非相互獨立（mutually independent）：若 $X_{12} = X_{23} = 1$，則必然 $X_{13} = 1$。計算方差只需要成對獨立，這個想法應用於隨機化算法（randomised algorithms）和生日問題（birthday problem）之中。第 (iii) 部分展示了當各項相依時，利用協方差計算和的方差的一般方法。兩者在 Blitzstein 與 Hwang 的 *Introduction to Probability*（《概率導論》）中均有詳細闡述。

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。