詳細解答

STEP III 2017，第 12 題

概率與統計 • 離散聯合分佈

> **STEP III 2017，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/17-s3-q12

關鍵思路

一切都源自兩條標準求和公式：$\sum_{r=1}^{n} r = \frac{1}{2}n(n+1)$ 及 $\sum_{r=1}^{n} r^2 = \frac{1}{6}n(n+1)(2n+1)$。首先令所有概率之和為 1，從而定出 k。然後對 y 求和，得出邊際分佈（marginal distribution）。至於協方差（covariance），我們需要 $\mathrm{E}(X)$ 和 $\mathrm{E}(XY)$。由於 $x+y$ 對 x 和 y 是對稱的，這種對稱性（symmetry）可以省去一半功夫。

解答

第 (i) 部分

所有 n^2 個數對 (x,y) 的概率之和必須為 1。由對稱性，$\sum_{x,y} x = \sum_{x,y} y$。由於對 y 的 n 個取值，每個 x 都各出現一次，所以兩者皆等於 $n\sum_{x=1}^{n} x = \frac{1}{2}n^2(n+1)$。因此

$$1 = k\sum_{x=1}^{n}\sum_{y=1}^{n}(x+y) = k\cdot 2\cdot \tfrac{1}{2}n^2(n+1) = k\,n^2(n+1), \qquad \text{故} \qquad k = \frac{1}{n^2(n+1)}.$$

求 X 的邊際分佈時，固定 x 並對 y 求和：

$$\begin{aligned}\mathrm{P}(X=x) &= k\sum_{y=1}^{n}(x+y) = k\Big(nx + \tfrac{1}{2}n(n+1)\Big) = \frac{kn\,(2x+n+1)}{2}\\ &= \frac{n+1+2x}{2n(n+1)},\end{aligned}$$

這正是所求結果。由於 $x+y$ 對 x 和 y 對稱，Y 具有相同的分佈：$\mathrm{P}(Y=y) = \dfrac{n+1+2y}{2n(n+1)}$。

想一想 要證明 X 與 Y 不獨立，是否需要逐一檢查每個數對 (x,y)？

不需要。獨立（independence）要求對每一個數對都有 $\mathrm{P}(X=x, Y=y) = \mathrm{P}(X=x)\,\mathrm{P}(Y=y)$，所以只要找到一個不成立的數對便足夠。取 $x=y=1$：

$$\mathrm{P}(X=1,Y=1) = 2k = \frac{2}{n^2(n+1)}, \qquad \mathrm{P}(X=1)\,\mathrm{P}(Y=1) = \frac{(n+3)^2}{4n^2(n+1)^2}.$$

兩者相等當且僅當 $8(n+1) = (n+3)^2$，即 $n^2 - 2n + 1 = 0$，亦即 $(n-1)^2 = 0$。所以兩者只在 $n=1$ 時相等。由於 $n \geq 2$，兩個概率不相等，因此 X 與 Y **不獨立**。

第 (ii) 部分

我們使用 $\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y)$，並留意由於兩個變量分佈相同，$\mathrm{E}(Y) = \mathrm{E}(X)$。

期望值（*expectation*）。利用第 (i) 部分的邊際分佈，

$$\begin{aligned}\mathrm{E}(X) &= \frac{1}{2n(n+1)} \sum_{x=1}^{n} ((n+1)x + 2x^2) = \frac{1}{2n(n+1)} \Big(\tfrac{1}{2} n(n+1)^2 + \tfrac{1}{3} n(n+1)(2n+1) \Big) \\ &= \frac{1}{2} \Big(\frac{n+1}{2} + \frac{2n+1}{3} \Big) = \frac{7n+5}{12}.\end{aligned}$$

乘積的期望值。這裏 $xy(x+y) = x^2y + xy^2$，由對稱性，兩部分的二重和相等。此外，x^2y 的二重和可分解為兩個單重和的乘積：

$$\begin{aligned}\mathrm{E}(XY) &= k \sum_{x=1}^{n} \sum_{y=1}^{n} xy(x+y) = 2k \Big(\sum_{x=1}^{n} x^2 \Big) \Big(\sum_{y=1}^{n} y \Big) = 2k \cdot \frac{n(n+1)(2n+1)}{6} \cdot \frac{n(n+1)}{2} \\ &= \frac{n^2(n+1)^2(2n+1)}{6\,n^2(n+1)} = \frac{(n+1)(2n+1)}{6}.\end{aligned}$$

協方差。以 144 為公分母，把結果合併：

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \frac{(n+1)(2n+1)}{6} - \frac{(7n+5)^2}{144} = \frac{24(2n^2+3n+1) - (49n^2+70n+25)}{144} \\ &= \frac{-n^2+2n-1}{144} = -\frac{(n-1)^2}{144}.\end{aligned}$$

由於 $n \geq 2$，$(n-1)^2 > 0$，所以 $\mathrm{Cov}(X,Y) = -\dfrac{(n-1)^2}{144} < 0$。協方差為負，正如所求。

協方差為負是合理的。聯合概率只取決於 $x+y$，傾向於使總和較大。如果 X 已經很大，Y 便不必同樣很大也能使總和變大，所以 Y 變大的「理由」稍為少了一點。

驗算

- 取 $n=2$。此時 $k=\frac{1}{12}$，四個概率分別為 $\frac{2}{12},\frac{3}{12},\frac{3}{12},\frac{4}{12}$，總和為 1。邊際概率為 $\mathrm{P}(X=1)=\frac{5}{12}$ 及 $\mathrm{P}(X=2)=\frac{7}{12}$，與公式吻合。於是 $\mathrm{E}(X)=\frac{19}{12}$，而 $\mathrm{E}(XY)=\frac{2+6+6+16}{12}=\frac{5}{2}$。由此得 $\mathrm{Cov}(X,Y)=\frac{5}{2}-\frac{361}{144}=-\frac{1}{144}$，與 $-(n-1)^2/144$ 一致。
- 邊際概率之和為 1：$\sum_x(n+1+2x)=n(n+1)+n(n+1)=2n(n+1)$。
- 令獨立性檢驗失效的 $(n-1)^2$，正正也是令協方差為零的那個值。在兩者中，$n=1$ 都是退化情況，此時必然有 $X=Y=1$。
- 以電腦對 $n=2,\ldots,29$ 的精確分佈作檢驗，確認了 k、邊際分佈、$\mathrm{E}(X)$、$\mathrm{E}(XY)$ 及協方差公式。

常見錯誤

- 忘記 $\sum_{x,y}x$ 中由對 y 求和而來的因子 n。這樣會得出 $k=1/(n(n+1))$，這是錯誤的。
- 試圖「一般地」證明不獨立，而不是給出一個明確的數對。單一個反例（counterexample）既已足夠，也是最清晰的論證。
- 以為必須有非零協方差才能證明相依。在本題中非零協方差確實顯示了相依，但協方差為零並不能證明獨立。

STEP 之外

形如 $f(x,y)=g(x+y)$ 的聯合分佈（joint distribution），是一個簡單的模型，描述通過總和而互相關聯的變量。此時協方差的正負取決於 g 的形狀。本題中 g 是線性的，協方差是一個很小的負數。在大學裏，聯合分佈、邊際分佈、條件分佈（conditional distribution）及相關（correlation）會一併學習，而連續情況則以二重積分代替二重和。Grimmett 與 Welsh 的 *Probability: An Introduction*（《概率導論》）以接近 STEP 的風格講解這些內容。

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。