详细解答

STEP III 2017，第 12 题

概率与统计 • 离散联合分布

STEP III 2017，第 12 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/17-s3-q12

关键思路

一切都源自两条标准求和公式：$\sum_{r=1}^{n} r = \frac{1}{2}n(n+1)$ 及 $\sum_{r=1}^{n} r^2 = \frac{1}{6}n(n+1)(2n+1)$。首先令所有概率之和为 1，从而定出 k。然后对 y 求和，得出边际分布（marginal distribution）。至于协方差（covariance），我们需要 $\mathrm{E}(X)$ 和 $\mathrm{E}(XY)$。由于 $x+y$ 对 x 和 y 是对称的，这种对称性（symmetry）可以省去一半功夫。

解答

第 (i) 部分

所有 n^2 个数对 (x,y) 的概率之和必须为 1。由对称性，$\sum_{x,y} x = \sum_{x,y} y$。由于对 y 的 n 个取值，每个 x 都各出现一次，所以两者皆等于 $n\sum_{x=1}^{n} x = \frac{1}{2}n^2(n+1)$。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 的边际分布时，固定 x 并对 y 求和：

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

这正是所求结果。由于 $x+y$ 对 x 和 y 对称，Y 具有相同的分布：$\mathrm{P}(Y=y) = \dfrac{n+1+2y}{2n(n+1)}$。

? **想一想** 要证明 X 与 Y 不独立，是否需要逐一检查每个数对 (x,y)？

不需要。独立（independence）要求对每一个数对都有 $\mathrm{P}(X=x, Y=y) = \mathrm{P}(X=x)\,\mathrm{P}(Y=y)$，所以只要找到一个不成立的数对便足够。取 $x=y=1$：

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

两者相等当且仅当 $8(n+1) = (n+3)^2$，即 $n^2 - 2n + 1 = 0$，亦即 $(n-1)^2 = 0$。所以两者只在 $n=1$ 时相等。由于 $n \geq 2$，两个概率不相等，因此 X 与 Y **不独立**。

第 (ii) 部分

我们使用 $\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y)$，并留意由于两个变量分布相同，$\mathrm{E}(Y) = \mathrm{E}(X)$。

期望值（expectation）。利用第 (i) 部分的边际分布，

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

乘积的期望值。这里 $xy(x+y) = x^2y + xy^2$，由对称性，两部分的二重和相等。此外，x^2y 的二重和可分解为两个单重和的乘积：

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

协方差。以 144 为公分母，把结果合并：

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

由于 $n \geq 2$，$(n-1)^2 > 0$，所以 $\mathrm{Cov}(X,Y) = -\dfrac{(n-1)^2}{144} < 0$。协方差为负，正如所求。协方差为负是合理的。联合概率只取决于 $x+y$，倾向于使总和较大。如果 X 已经很大，Y 便不必同样很大也能使总和变大，所以 Y 变大的“理由”稍为少了一点。

验算

- 取 $n=2$。此时 $k = \frac{1}{12}$，四个概率分别为 $\frac{2}{12}, \frac{3}{12}, \frac{3}{12}, \frac{4}{12}$，总和为 1。边际概率为 $\mathrm{P}(X=1) = \frac{5}{12}$ 及 $\mathrm{P}(X=2) = \frac{7}{12}$，与公式吻合。于是 $\mathrm{E}(X) = \frac{19}{12}$，而 $\mathrm{E}(XY) = \frac{2+6+6+16}{12} = \frac{5}{2}$。由此得 $\mathrm{Cov}(X,Y) = \frac{5}{2} - \frac{361}{144} = -\frac{1}{144}$，与 $-(n-1)^2/144$ 一致。

- 边际概率之和为 1：$\sum_x (n+1+2x) = n(n+1) + n(n+1) = 2n(n+1)$。
- 令独立性检验失效的 $(n-1)^2$，正正也是令协方差为零的那个值。在两者中，$n=1$ 都是退化情况，此时必然有 $X = Y = 1$。
- 以电脑对 $n = 2, \ldots, 29$ 的精确分布作检验，确认了 k、边际分布、$\mathrm{E}(X)$、$\mathrm{E}(XY)$ 及协方差公式。

常见错误

- 忘记 $\sum_{x,y} x$ 中由对 y 求和而来的因子 n。这样会得出 $k = 1/(n(n+1))$，这是错误的。
- 试图“一般地”证明不独立，而不是给出一个明确的数对。单一个反例（counterexample）既已足够，也是最清晰的论证。
- 以为必须有非零协方差才能证明相依。在本题中非零协方差确实显示了相依，但协方差为零并不能证明独立。

STEP 之外

形如 $f(x,y) = g(x+y)$ 的联合分布（joint distribution），是一个简单的模型，描述通过总和而互相关联的变量。此时协方差的正负取决于 g 的形状。本题中 g 是线性的，协方差是一个很小的负数。在大学里，联合分布、边际分布、条件分布（conditional distribution）及相关（correlation）会一并学习，而连续情况则以二重积分代替二重和。Grimmett 与 Welsh 的 *Probability: An Introduction*（《概率导论》）以接近 STEP 的风格讲解这些内容。

Beyond STEP 详细解答：https://beyondstep.keithmcnulty.org/zh-hans/。解答文字采用 CC0 1.0；题目本身版权属 Cambridge University Press & Assessment 所有（©），本文件不予转载。由机器从英文翻译而成；尚未经审阅。