解答例

STEP III 2017、第 12 問

確率・統計 ・ 離散型同時分布

> **STEP III 2017、問 12**
>
> 先に進む前に、まず問題を読んでおこう。問題は STEP Support Programme の問題データベースにある：
>
> step.maths.org/questions/17-s3-q12

鍵となる考え方

すべては二つの標準的な和の公式 $\sum_{r=1}^{n} r = \frac{1}{2}n(n+1)$ と $\sum_{r=1}^{n} r^2 = \frac{1}{6}n(n+1)(2n+1)$ から導かれる。まず、確率（probability）の総和が 1 になることから k を定める。次に y について和をとることで周辺分布（marginal distribution）を求める。共分散（covariance）を求めるには $\mathrm{E}(X)$ と $\mathrm{E}(XY)$ が必要である。$x+y$ が x と y について対称（symmetric）であることを使うと、計算量は半分になる。

解答

小問 (i)

n^2 個のすべての組 (x,y) にわたる確率の総和は 1 でなければならない。対称性より $\sum_{x,y} x = \sum_{x,y} y$ である。各 x は y の n 個の値それぞれに対して一度ずつ現れるので、これらはいずれも $n\sum_{x=1}^{n} x = \frac{1}{2}n^2(n+1)$ に等しい。したがって

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

X の周辺分布を求めるには、x を固定して y について和をとる：

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

これが求める結果である。$x+y$ は x と y について対称なので、Y も同じ分布をもつ：$\mathrm{P}(Y=y) = \dfrac{n+1+2y}{2n(n+1)}$。

考えてみよう X と Y が独立でないことを示すには、すべての組 (x,y) を調べる必要があるだろうか？

その必要はない。独立（independent）であるためには**すべての**組について $\mathrm{P}(X=x, Y=y) = \mathrm{P}(X=x)\,\mathrm{P}(Y=y)$ が成り立たなければならないので、これが成り立たない組が一つあれば十分である。$x=y=1$ とすると

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

これらが等しくなるのは $8(n+1) = (n+3)^2$、すなわち $n^2-2n+1=0$、つまり $(n-1)^2=0$ のときに限る。よって等しくなるのは $n=1$ のときだけである。$n \geq 2$ なので二つの確率は異なり、X と Y は**独立でない**。

小問 (ii)

$\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y)$ を用いる。また、二つの確率変数は同じ分布をもつので $\mathrm{E}(Y) = \mathrm{E}(X)$ であることに注意する。

平均。小問 (i) の周辺分布を用いると

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

積の平均。ここで $xy(x+y) = x^2y + xy^2$ であり、対称性より二つの項の二重和は等しい。さらに、x^2y の二重和は一重の和の積に分解できる：

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

共分散。以上を共通の分母 144 でまとめると

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

$n \geq 2$ より $(n-1)^2 > 0$ なので、$\mathrm{Cov}(X,Y) = -\dfrac{(n-1)^2}{144} < 0$ である。よって共分散は負であり、題意は示された。

共分散が負になるのは自然なことである。同時確率は $x+y$ だけで決まり、合計が大きいほど起こりやすい。X がすでに大きければ、合計を大きくするために Y まで大きくなる必要はないので、Y が大きくなる理由はわずかに弱まるのである。

検算

- $n=2$ とする。このとき $k=\frac{1}{12}$ であり、四つの確率は $\frac{2}{12}, \frac{3}{12}, \frac{3}{12}, \frac{4}{12}$ で、総和は 1 になる。周辺分布は $\mathrm{P}(X=1)=\frac{5}{12}$、$\mathrm{P}(X=2)=\frac{7}{12}$ で、公式と一致する。したがって $\mathrm{E}(X)=\frac{19}{12}$、$\mathrm{E}(XY)=\frac{2+6+6+16}{12}=\frac{5}{2}$ である。これより $\mathrm{Cov}(X,Y)=\frac{5}{2}-\frac{361}{144}=-\frac{1}{144}$ となり、$-(n-1)^2/144$ と一致する。
- 周辺確率の総和は 1 になる：$\sum_x (n+1+2x)=n(n+1)+n(n+1)=2n(n+1)$。
- 独立性の判定を破綻させる $(n-1)^2$ の値は、共分散を 0 にする値と同じである。どちらにおいても、$n=1$ は確率 1 で $X=Y=1$ となる退化した場合である。
- $n=2,\ldots,29$ について厳密な分布をコンピュータで確認し、k、周辺分布、$\mathrm{E}(X)$、$\mathrm{E}(XY)$、共分散の公式が正しいことを確かめた。

よくある間違い

- $\sum_{x,y} x$ には y についての和から因数 n が掛かることを忘れる。この場合 $k=1/(n(n+1))$ となり、誤りである。
- 具体的な組を一つ示すのではなく、「一般的に」独立でないことを示そうとする。反例（counterexample）は一つで十分であり、しかも最も明快な論証である。
- 従属性を示すには共分散が 0 でないことが必要だと考える。ここでは確かに共分散によって従属性が示されるが、共分散が 0 であっても独立性が証明されるわけではない。

STEP のその先へ

$f(x,y)=g(x+y)$ という形の同時分布（joint distribution）は、合計を通じて結びついた確率変数の簡単なモデルである。このとき共分散の符号は g の形によって決まる。ここでは g は一次関数であり、共分散は小さな負の値になる。大学では、同時分布、周辺分布、条件付き分布（conditional distribution）、相関（correlation）をまとめて学び、連続型の場合には二重和（double sum）の代わりに二重積分（double integral）を用いる。Grimmett と Welsh による *Probability: An Introduction*（確率論入門）は、STEP に近いスタイルでこの内容を扱っている。

Beyond STEP 解答集：https://beyondstep.keithmcnulty.org/ja/。解答本文は CC0 1.0 である。問題そのものは © Cambridge University Press & Assessment に帰属し、ここには掲載していない。英語版から機械翻訳したものであり、まだ校閲を経ていない。