解答例

STEP III 2000、第 14 問

確率・統計 ・ 共分散と独立性

> **STEP III 2000、問 14**
>
> 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/00-s3-q14

鍵となる考え方

欠けている 4 つ目の確率を埋めてから、$\mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$ を項ごとに書き出す。すべては 1 つの因数 $a - pq$ と、決して 0 にならない式との積にまとまる。そして $a = pq$ とは、まさに左上のセルが積に分解される（因数分解できる）という主張にほかならない。2×2 の表では、1 つのセルが積に分解されれば、周辺の和の条件によって残りの 3 つのセルも同様に分解される。値が 3 つになると自由度が増え、対称な表によってこの結び付きが断ち切られる。

解答

恒等式

4 つの確率の和は 1 であるから、

$$\mathrm{P}(X = x_2, Y = y_2) = 1 - a - (q - a) - (p - a) = 1 - p - q + a.$$

行と列に沿って足し合わせると、周辺分布（marginal distribution）が得られる：

$$\mathrm{P}(X = x_1) = a + (q - a) = q, \qquad \mathrm{P}(Y = y_1) = a + (p - a) = p,$$

したがって $\mathrm{P}(X = x_2) = 1 - q$、$\mathrm{P}(Y = y_2) = 1 - p$ である。これが p と q の本当の意味である。

	y_1	y_2	
x_1	a	$q-a$	q
x_2	$p-a$	$1-p-q+a$	$1-q$
	p	$1-p$	

ここで

$$\begin{aligned}\mathrm{E}(XY) &= a\,x_1y_1 + (q-a)\,x_1y_2 + (p-a)\,x_2y_1 + (1-p-q+a)\,x_2y_2,\\ \mathrm{E}(X)\mathrm{E}(Y) &= \big(qx_1 + (1-q)x_2\big)\big(py_1 + (1-p)y_2\big)\\ &= pq\,x_1y_1 + q(1-p)\,x_1y_2 + p(1-q)\,x_2y_1 + (1-p)(1-q)\,x_2y_2.\end{aligned}$$

係数ごとに引き算すると、

$$\begin{aligned}x_1y_1 :&\quad a - pq,\\ x_1y_2 :&\quad (q-a) - (q-pq) = -(a-pq),\\ x_2y_1 :&\quad (p-a) - (p-pq) = -(a-pq),\\ x_2y_2 :&\quad (1-p-q+a) - (1-p-q+pq) = a-pq.\end{aligned}$$

よって

$$\mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) = (a-pq)(x_1y_1 - x_1y_2 - x_2y_1 + x_2y_2) = (a-pq)(x_1-x_2)(y_1-y_2).$$

したがって $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$ ならば左辺は 0 であり、$(a-pq)(x_1-x_2)(y_1-y_2) = 0$ となる。これが示すべきことであった。

独立性

それぞれが相異なる 2 つの値をとる確率変数（random variable）の組は、適当な a, p, q を用いて上の形の同時分布表（joint table）をもつ（各セルに名前を付ければよい）。$\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$ ならば上の積は 0 である。$x_1 \neq x_2$ かつ $y_1 \neq y_2$ であるから後ろの 2 つの因数は 0 でなく、ゆえに $a = pq$ である。

> **考えてみよう** これで 1 つのセルが積に分解されることがわかった。なぜそれで残りの 3 つも決まるのだろうか？

それは、他の各セルが「周辺確率から既知のセルを引いたもの」になっているからで

ある：

$$\begin{aligned}
&\mathrm{P}(X=x_1, Y=y_2) = q - pq = q(1-p) = \mathrm{P}(X=x_1)\,\mathrm{P}(Y=y_2),\\
&\mathrm{P}(X=x_2, Y=y_1) = p - pq = p(1-q) = \mathrm{P}(X=x_2)\,\mathrm{P}(Y=y_1),\\
&\mathrm{P}(X=x_2, Y=y_2) = 1 - p - q + pq = (1-p)(1-q) = \mathrm{P}(X=x_2)\,\mathrm{P}(Y=y_2).
\end{aligned}$$

すべてのセルが周辺確率の積になっているので、X と Y は独立（independent）である。（逆に、独立ならば $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$ が成り立つことはよく知られているので、2 値の確率変数についてはこの 2 つの条件は同値である。）

3 値の例

A と B はいずれも $\{-1, 0, 1\}$ 上の一様分布（uniform distribution）に従うから $\mathrm{E}(A) = \mathrm{E}(B) = 0$ である。したがって、どの行の和もどの列の和も $\frac{1}{3}$ で、$\mathrm{E}(AB) = 0$ となるが、すべてのセルが $\frac{1}{9}$ に等しいわけではないような表を作ればよい。対称な選び方がうまくいく：

	$B=-1$	$B=0$	$B=1$
$A=-1$	$\frac{1}{6}$	0	$\frac{1}{6}$
$A=0$	0	$\frac{1}{3}$	0
$A=1$	$\frac{1}{6}$	0	$\frac{1}{6}$

各行・各列の和は $\frac{1}{3}$ であるから、A と B はそれぞれ $-1, 0, 1$ を確率 $\frac{1}{3}$ でとる。積 ab が 0 でないのは四隅のみであり、

$$\mathrm{E}(AB) = \tfrac{1}{6}(1) + \tfrac{1}{6}(-1) + \tfrac{1}{6}(-1) + \tfrac{1}{6}(1) = 0 = \mathrm{E}(A)\mathrm{E}(B).$$

しかし $\mathrm{P}(A=0, B=0) = \frac{1}{3} \neq \frac{1}{9} = \mathrm{P}(A=0)\,\mathrm{P}(B=0)$ であるから、A と B は独立ではない。言葉で言えば、$B=0$ となるのはちょうど $A=0$ のときであり、それ以外のときは A の値によらず B は等確率で ± 1 をとる。

検算

- $X = Y$、$x_1 = y_1 = 1$、$x_2 = y_2 = 0$ としてみる。このとき $a = p = q$ であり、恒等式は $\mathrm{Var}(X) = p - p^2$ を与える。これはベルヌーイ（Bernoulli）変数として正しい。
- この恒等式は多項式の恒等式として成り立つ。x_i, y_i, a, p, q を無作為に 2000 通り選び、有理数演算で厳密に確かめた。同じ計算で 3 値の表も確認した：周辺確率はすべて $\frac{1}{3}$、$\mathrm{E}(AB) = 0$、$\mathrm{P}(A=0, B=0) = \frac{1}{3}$ である。

よくある間違い

- $a = pq$ を示したところで止めてしまう。独立性には**すべての**セルが積に分解されることが必要なので、残りの 3 つも（手短にでよいから）確かめる必要がある。
- $(x_1 - x_2)(y_1 - y_2)$ で割ってよい理由を書き忘れる。値が相異なるので、これは 0 でない。
- 最後の小問で $B = A^2$ や $B = |A|$ を答えてしまう。これらは共分散が 0 で独立でない例ではあるが、B が -1 をとることがないので、要求された周辺分布をもたない。

STEP のその先へ

量 $\mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$ は共分散（covariance）であり、この問題は「無相関（uncorrelated）」が「独立」よりも弱い条件であることを示している。ただし 2 値の（指示変数（indicator）のような）確率変数は例外であり、その場合は恒等式によってすべてが 2 つの指示変数の共分散に帰着する。今後出会う関連した事実として、同時正規分布（jointly normal）に従う確率変数については、共分散が 0 ならば独立であることが成り立つ。さらに学ぶには、Grimmett と Welsh の *Probability: An Introduction*（確率論入門）や、Blitzstein と Hwang の *Introduction to Probability*（確率入門）を参照するとよい。

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。