解答例

STEP III 2002、第 14 問

確率・統計 ・ 分散と共分散

> **STEP III 2002、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/02-s3-q14

鍵となる考え方

分散（variance）の恒等式は、分散の定義を期待値の線形性（linearity of expectation）を用いて展開したものにすぎない。チェスの部分は見た目以上に巧妙である。$\mathrm{E}(WL)$ を直接計算することは一度もない。その代わりに、$W+L$ の分散は二項分布（binomial distribution）からわかり、$\mathrm{Var}(W)$ と $\mathrm{Var}(L)$ も二項分布からわかる。そして、いま証明したばかりの恒等式が共分散（covariance）をただで与えてくれるのである。

解答

分散の恒等式

$\mu_X = \mathrm{E}(X)$、$\mu_Y = \mathrm{E}(Y)$ とおく。次の定義

$$\mathrm{Var}(X) = \mathrm{E}[(X-\mu_X)^2], \qquad \mathrm{Cov}(X,Y) = \mathrm{E}[(X-\mu_X)(Y-\mu_Y)],$$

および期待値の線形性 $\mathrm{E}(aU+bV) = a\,\mathrm{E}(U) + b\,\mathrm{E}(V)$ を用いる。線形性は、離散確率変数（discrete random variable）の場合、和 $\sum (au+bv)\,\mathrm{P}(U=u, V=v)$ を二つの和に分け、それぞれでもう一方の変数について和をとってしまえば得られる。

線形性より $\mathrm{E}(X+Y) = \mu_X + \mu_Y$ であるから、

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathrm{E}[((X-\mu_X)+(Y-\mu_Y))^2] \\
&= \mathrm{E}[(X-\mu_X)^2 + (Y-\mu_Y)^2 + 2(X-\mu_X)(Y-\mu_Y)] \\
&= \mathrm{E}[(X-\mu_X)^2] + \mathrm{E}[(Y-\mu_Y)^2] + 2\,\mathrm{E}[(X-\mu_X)(Y-\mu_Y)] \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y),
\end{aligned}$$

となる。ここで 3 行目では再び線形性を用いた。これが求める結果である。独立性（independence）はまったく必要なかったことに注意せよ。この恒等式は任意の X と Y について成り立つ。

小問 (i)

どの対局も勝ち、負け、引き分けのいずれかであるから、何が起ころうと $W+L+D=m$ である。常に同じ値をとる確率変数はその平均に等しいので、その分散は $\mathrm{E}[(m-m)^2]=0$ である。

小問 (ii)

勝ちまたは負けで終わった対局を**決着局**と呼ぶことにする。各対局が決着局となる確率は $\frac{1}{3}+\frac{1}{3}=\frac{2}{3}$ であり、対局どうしは独立で、対局数は一定の m である。したがって、決着局の数 $W+L$ は、成功確率の等しい m 回の独立な試行における成功回数を数えたものであり、

$$W+L \sim \mathrm{B}(m, \tfrac{2}{3}).$$

相関係数

まったく同様に、W はそれぞれ確率 $\frac{1}{3}$ で勝ちとなる m 回の独立な試行における勝ちの回数を数えたものであるから $W \sim \mathrm{B}(m, \frac{1}{3})$ であり、同様に $L \sim \mathrm{B}(m, \frac{1}{3})$ である。二項分布の分散の公式 $\mathrm{Var}=np(1-p)$ を用いると、

$$\mathrm{Var}(W)=\mathrm{Var}(L)=m\cdot\tfrac{1}{3}\cdot\tfrac{2}{3}=\frac{2m}{9}, \qquad \mathrm{Var}(W+L)=m\cdot\tfrac{2}{3}\cdot\tfrac{1}{3}=\frac{2m}{9}.$$

? **考えてみよう** $\mathrm{E}(WL)$ を計算せずに、どうすれば $\mathrm{Cov}(W,L)$ を求められるだろうか。

恒等式を逆向きに使えばよい。$X=W$、$Y=L$ とすると、

$$\frac{2m}{9}=\frac{2m}{9}+\frac{2m}{9}+2\,\mathrm{Cov}(W,L), \qquad \text{よって} \qquad \mathrm{Cov}(W,L)=-\frac{m}{9}.$$

したがって、$m \geq 1$（分散が 0 でないため）のとき、相関係数（correlation）は

$$\frac{\mathrm{Cov}(W,L)}{\sqrt{\mathrm{Var}(W)\,\mathrm{Var}(L)}}=\frac{-m/9}{2m/9}=-\frac{1}{2}.$$

答えは m によらない。また負であるのは当然である。対局数が一定なので、勝ちが多いほど負けの入る余地は少なくなる。

小問 (i) からは別の道筋も得られる。恒等式を 2 回適用すると、$\mathrm{Var}(W+L+D)$ は三つの分散の和に、三つの組ごとの共分散の和の 2 倍を加えたものになる。対称性より

三つの共分散はすべてある値 c に等しいので、$0 = 3 \cdot \frac{2m}{9} + 6c$ となり、やはり $c = -\frac{m}{9}$ を得る。

検算

- $m = 1$ とする。このとき W と L は指示変数（indicator）で、$\mathrm{E}(W) = \mathrm{E}(L) = \frac{1}{3}$ かつ常に $WL = 0$ であるから、$\mathrm{Cov}(W, L) = 0 - \frac{1}{9} = -\frac{1}{9}$ となり、$-m/9$ と一致する。
- 二つの道筋（小問 (ii) を経由するものと小問 (i) を経由するもの）の結果は一致する。
- $m = 1, \ldots, 7$ について 3^m 通りの結果をすべて厳密に数え上げると、いずれの場合も $\mathrm{Var}(W) = 2m/9$、$\mathrm{Cov}(W, L) = -m/9$ であり、相関係数はちょうど $-\frac{1}{2}$ となる。

よくある間違い

- **対局**が独立だから W と L も独立だと思い込むこと。実際には独立ではない。D とともに一定の合計を分け合っているからである。
- $\mathrm{Var}(W + L) = \mathrm{Var}(W) + \mathrm{Var}(L) = 4m/9$ としてしまうこと。これは共分散の項を忘れており、二項分布から得られる答え $2m/9$ と矛盾する。

STEP のその先へ

(W, L, D) は**多項分布**（multinomial distribution）に従う。これは、1 回の試行の結果が 3 通り以上ある場合への二項分布の自然な一般化である。一般に、各結果の確率が $p_1, \ldots, p_k$ のとき、回数 N_i と N_j の共分散は $-mp_ip_j$、相関係数は $-\sqrt{p_ip_j/\big((1-p_i)(1-p_j)\big)}$ であり、ここではこれが $-\frac{1}{2}$ を与える。共分散行列（covariance matrix）や共分散の双線形性（bilinearity）については、Grimmett と Welsh による *Probability: An Introduction*（確率論入門）や、Blitzstein と Hwang による *Introduction to Probability*（確率論入門）で詳しく論じられている。

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。