詳細解答

STEP III 2002，第 14 題

概率與統計 • 方差與協方差

> **STEP III 2002，第 14 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/02-s3-q14

關鍵思路

這條方差恆等式其實就是方差（variance）的定義，再利用期望值的線性性質（linearity of expectation）展開而已。國際象棋那部分比表面看來更巧妙：我們從頭到尾都沒有直接計算 $\mathrm{E}(WL)$。相反，我們由二項分佈（binomial distribution）得知 $W+L$ 的方差，同樣由二項分佈得知 $\mathrm{Var}(W)$ 和 $\mathrm{Var}(L)$，而剛證明的恆等式便把協方差（covariance）白白送到我們手上。

解答

方差恆等式

記 $\mu_X = \mathrm{E}(X)$ 及 $\mu_Y = \mathrm{E}(Y)$。我們採用以下定義

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

以及期望值的線性性質 $\mathrm{E}(aU+bV) = a\,\mathrm{E}(U) + b\,\mathrm{E}(V)$；對於離散隨機變量（discrete random variable），只要把和式 $\sum (au+bv)\,\mathrm{P}(U=u, V=v)$ 拆成兩個和式，並在每個和式中對另一個變量求和，便可得出這個性質。

根據線性性質，$\mathrm{E}(X+Y) = \mu_X + \mu_Y$，所以

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

其中第三行再次用到線性性質。這就是所需的結果。留意我們並不需要獨立性（independence）：這條恆等式對任何 X 和 Y 都成立。

第 (i) 部分

每一局棋不是勝、就是負、或者和，所以無論結果如何，都有 $W + L + D = m$。一個永遠取同一個值的隨機變量等於它的平均值，所以它的方差是 $\mathrm{E}[(m-m)^2] = 0$。

第 (ii) 部分

如果一局棋分出了勝負（即勝或負），我們稱它為分勝負的一局。每一局分勝負的概率是 $\frac{1}{3} + \frac{1}{3} = \frac{2}{3}$，各局互相獨立，而且局數固定為 m。因此 $W + L$，即分勝負的局數，就是在 m 次成功概率相同的獨立試驗中成功的次數：

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

相關係數

同樣道理，W 是 m 次獨立試驗中勝出的次數，每次勝出的概率是 $\frac{1}{3}$，所以 $W \sim \mathrm{B}(m, \frac{1}{3})$，同理 $L \sim \mathrm{B}(m, \frac{1}{3})$。利用二項分佈的公式 $\mathrm{Var} = np(1-p)$，

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

想一想 怎樣可以不計算 $\mathrm{E}(WL)$ 而求出 $\mathrm{Cov}(W, L)$？

把恆等式倒過來用。取 $X = W$ 及 $Y = L$，

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

因此，當 $m \geq 1$ 時（這樣方差才不是零），

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

答案與 m 無關。它是負數，這亦合乎情理：局數既然固定，勝的局數越多，留給負局的空間便越少。

第 (i) 部分提供了另一條途徑。把恆等式應用兩次，可知 $\mathrm{Var}(W + L + D)$ 等於三個方差之和，再加上三個兩兩協方差之和的兩倍。根據對稱性，三個協方差全部相等，設為 c，於是 $0 = 3 \cdot \frac{2m}{9} + 6c$，再次得出 $c = -\frac{m}{9}$。

驗算

- 取 $m = 1$。這時 W 和 L 都是指示變量（indicator），$\mathrm{E}(W) = \mathrm{E}(L) = \frac{1}{3}$，而且恆有 $WL = 0$，所以 $\mathrm{Cov}(W, L) = 0 - \frac{1}{9} = -\frac{1}{9}$，與 $-m/9$ 吻合。
- 兩條途徑（經第 (ii) 部分和經第 (i) 部分）的結果一致。
- 對 $m = 1, \ldots, 7$ 逐一窮舉全部 3^m 個結果，每次都得出 $\mathrm{Var}(W) = 2m/9$、$\mathrm{Cov}(W, L) = -m/9$，而相關係數（correlation）都恰好是 $-\frac{1}{2}$。

常見錯誤

- 因為各局棋互相獨立，便以為 W 和 L 也互相獨立。其實不然：它們與 D 合起來的總數是固定的。
- 取 $\mathrm{Var}(W + L) = \mathrm{Var}(W) + \mathrm{Var}(L) = 4m/9$，這樣做漏掉了協方差一項，而且與二項分佈得出的答案 $2m/9$ 矛盾。

STEP 之外

(W, L, D) 服從*多項分佈*（multinomial distribution），即把二項分佈自然地推廣至每次試驗有多於兩種結果的情況。一般而言，若各種結果的概率為 $p_1, \ldots, p_k$，則計數 N_i 與 N_j 的協方差為 $-mp_ip_j$，相關係數為 $-\sqrt{p_ip_j/\big((1-p_i)(1-p_j)\big)}$，在本題中就是 $-\frac{1}{2}$。協方差矩陣（covariance matrix）及協方差的雙線性（bilinearity），可參閱 Grimmett 與 Welsh 的 *Probability: An Introduction*（概率論入門），以及 Blitzstein 與 Hwang 的 *Introduction to Probability*（概率論導論）。

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。