詳細解答

STEP I 1988，第 15 題

概率與統計 • 二項分佈的假設檢驗

> **STEP I 1988，第 15 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/88-s1-q15

關鍵思路

把「超過一半的得分來自犯規」化為犯規次數與入球數之間的一個簡單不等式（inequality）。以入球數作為條件（conditioning）之後，餘下的工作只是數一數有多少個等可能的犯規次數符合要求。至於第二部分，要問的是：按照舊分佈，在 300 次試驗中預期會出現多少次九次犯規的情況？假如甚麼都沒有改變，觀察到的次數又有多出人意料？

解答

概率

就某場比賽中的某一支球隊而言，設 G 為該隊的入球數，F 為對方向該隊犯規的次數。那麼 G 與 F 獨立（independent），且

$$\mathrm{P}(G=0),\ \ldots,\ \mathrm{P}(G=4) \;=\; \tfrac{1}{9},\ \tfrac{2}{9},\ \tfrac{3}{9},\ \tfrac{2}{9},\ \tfrac{1}{9}, \qquad \mathrm{P}(F=f)=\tfrac{1}{10} \quad (f=0,1,\ldots,9).$$

該隊從入球得 $2G$ 分，從犯規得 F 分，合共 $2G+F$ 分。該隊超過一半的得分來自犯規，當且僅當

$$F > \tfrac{1}{2}(2G+F) \iff F > 2G.$$

（假如該隊一分也沒有得到，則 $F = 0 = 2G$，這種情況不應計算在內，而嚴格不等式正好把它排除了。）

給定 $G = g$，符合要求的 F 值為 $2g+1,\ldots,9$，即十個等可能的值當中的 $9-2g$ 個。因此，利用 F 與 G 的獨立性，

$$\mathrm{P}(F > 2G \mid G=g) = \frac{9-2g}{10}, \qquad g=0,1,2,3,4,$$

得出 $\frac{9}{10}, \frac{7}{10}, \frac{5}{10}, \frac{3}{10}, \frac{1}{10}$。根據全概率公式（law of total probability），

$$\mathrm{P}(F > 2G) = \frac{1\cdot 9 + 2\cdot 7 + 3\cdot 5 + 2\cdot 3 + 1\cdot 1}{9\times 10} = \frac{45}{90} = \frac{1}{2}.$$

答案如此簡潔，是有原因的：入球數的分佈關於 2 對稱，而 $(9-2g)/10$ 是 g 的線性函數，所以它的平均值等於它在 $g=2$ 的值，即 $\frac{1}{2}$。

犯規次數的分佈有沒有改變？

一支球隊的犯規次數，就是對手被犯規的次數，所以在舊規則下，每支球隊在一場比賽中犯規 9 次的概率為 $\frac{1}{10}$。設 X 為 300 個球隊場次當中，球隊犯規 9 次的場次數目。我們檢驗

$$H_0:\ p = \tfrac{1}{10}\ \text{（沒有改變）} \qquad \text{對} \qquad H_1:\ p < \tfrac{1}{10},$$

其中 p 是一支球隊在一場比賽中犯規 9 次的概率。備擇假設（alternative hypothesis）是單側（one-sided）的，因為修改規則的目的，是令犯規次數偏高的情況變得較不可能出現。

? **想一想** 這裏採用二項分佈（binomial）模型是否合理？

我們需要這 300 個球隊場次互相獨立，而且每個場次的概率 p 都相同。題目已說明同一場比賽之內的獨立性，而假設不同場次的比賽之間也互相獨立，亦屬合理。因此在 H_0 之下，$X \sim \mathrm{B}(300, \frac{1}{10})$，且

$$\mathrm{E}(X) = 30, \qquad \mathrm{Var}(X) = 300\times\tfrac{1}{10}\times\tfrac{9}{10} = 27, \qquad \text{標準差}\sqrt{27}\approx 5.20.$$

我們觀察到 $X=12$，比期望值 30 少了 18，相當於約 3.5 個標準差（standard deviation）。採用正態近似（normal approximation）並加上連續性校正（continuity correction），

$$\mathrm{P}(X\le 12) \approx \Phi\left(\frac{12.5-30}{\sqrt{27}}\right) = \Phi(-3.37) \approx 0.0004.$$

精確的二項概率還要更小，約為 0.0001。無論採用哪一個數值，假如甚麼都沒有改變，這樣極端的結果在一千個球季之中也出現不到一次。它遠遠超出常用的 5% 甚至 1% 顯著性水平（significance level），所以我們拒絕 H_0。這*確實*是有力的證據，顯示犯規次數的分佈已經改變，而且是朝着所期望的方向改變：犯規九次的場次變得較為罕見。

有兩點保留值得說明。這個檢驗只顯示 P(9 次犯規) 下降了，並沒有說明分佈的其餘部分如何改變（儘管那是最自然的解釋）。此外，統計學也無法證明這個下降是由修改規則*造成*的，因為兩個球季之間可能還有其他事情改變了。

驗算

► 入球數的概率 $\frac{1}{9}(1+2+3+2+1)$ 之和為 1，理應如此。

- 以精確分數逐一列舉全部 5×10 個數對 (g, f)，得出 $\mathrm{P}(F > 2G) = \frac{1}{2}$；而模擬 200,000 場比賽的結果為 0.4997。
- 直接把二項概率加起來，對於 $X \sim \mathrm{B}(300, 0.1)$ 得出 $\mathrm{P}(X \leq 12) = 9.2 \times 10^{-5}$。正態近似（$3.8 \times 10^{-4}$）與平均值為 30 的泊松近似（Poisson approximation）（1.7×10^{-4}）都一致顯示這個概率極小。

常見錯誤

- 把 F 與 G 比較，而不是與 $2G$ 比較：一個入球值兩分。
- 計算 $F \geq 2G$ 而不是 $F > 2G$。「超過一半」是嚴格的不等關係，而一分未得的球隊，並沒有任何得分來自犯規。
- 以 150 場比賽作為樣本大小，而不是 300 個球隊場次；或者在計算之前忘記寫明零假設（null hypothesis）和二項分佈模型。

STEP 之外

第二部分是對二項比例所作的假設檢驗（hypothesis test），而我們求得的尾部概率就是 p 值（p-value）。到了大學，你會學到似然比（likelihood ratio）檢驗和卡方檢驗（chi-squared test），它們會把觀察所得的整個犯規次數分佈與舊分佈作比較，而不只是比較其中一個值。Blitzstein 與 Hwang 合著的 *Introduction to Probability*（概率論入門）把概率方面的內容（條件概率、二項分佈與正態近似）講解得十分清楚。

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。