詳細解答

STEP I 2008，第 12 題

概率與統計 • 離散均勻分佈變量的最大值

STEP I 2008，第 12 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/08-s1-q12

關鍵思路

X 其實就是兩個分數中較大的一個。若直接求「較大者等於 r」的概率，便要分幾種情況來數；但「較大者*不超過* r」卻很容易處理：這恰好在*兩個*分數都不超過 r 時發生。因此我們先求累積分佈（cumulative distribution）$\mathrm{P}(X \leq r)$，再取差。有了這個累積分佈，中位數（median）也幾乎立即可得。

解答

設 X_1 和 X_2 為兩個獨立（independent）的分數，並設 $X = \max(X_1, X_2)$（當 $X_1 = X_2$ 時，定義中的兩種情況給出相同的值，所以 X 是明確定義的）。

第 (i) 部分

對於 $r = 0, 1, \ldots, N$，事件 $\{X \leq r\}$ 就是事件 $\{X_1 \leq r$ 且 $X_2 \leq r\}$。這兩個事件各自的概率都是 r/N，而且由於 X_1 和 X_2 獨立，它們亦互相獨立，所以

$$\mathrm{P}(X \leq r) = \frac{r}{N} \cdot \frac{r}{N} = \frac{r^2}{N^2}.$$

這在 $r = 0$ 時同樣成立，所得的值為 0。因此，對於 $r = 1, 2, \ldots, N$，

$$\mathrm{P}(X = r) = \mathrm{P}(X \leq r) - \mathrm{P}(X \leq r - 1) = \frac{r^2 - (r-1)^2}{N^2} = \frac{2r - 1}{N^2},$$

這正是所要證明的結果。（簡單核對一下：$\sum_{r=1}^{N}(2r-1) = N^2$，所以這些概率之和為 1。）

第 (ii) 部分

利用題目給出的兩個求和公式，

$$\begin{aligned}\mu = \sum_{r=1}^{N} r\,\frac{2r-1}{N^2} &= \frac{1}{N^2}\left(2\sum_{r=1}^{N} r^2 - \sum_{r=1}^{N} r\right)\\ &= \frac{1}{N^2}\left(\frac{N(N+1)(2N+1)}{3} - \frac{N(N+1)}{2}\right)\\ &= \frac{N+1}{N}\cdot\frac{2(2N+1)-3}{6} = \frac{(N+1)(4N-1)}{6N}.\end{aligned}$$

當 $N = 100$ 時，

$$\mu = \frac{101\times 399}{600} = \frac{40\,299}{600} = 67.165,$$

與題目所求相符。

第 (iii) 部分

由第 (i) 部分，$\mathrm{P}(X \le m) = m^2/N^2$，而 $\mathrm{P}(X \ge m) = 1 - \mathrm{P}(X \le m-1) = 1 - (m-1)^2/N^2$。中位數須滿足的兩個條件於是變成

$$\mathrm{P}(X \le m) \ge \tfrac{1}{2} \iff m^2 \ge \tfrac{1}{2}N^2 \iff m \ge \frac{N}{\sqrt{2}},$$
$$\mathrm{P}(X \ge m) \ge \tfrac{1}{2} \iff (m-1)^2 \le \tfrac{1}{2}N^2 \iff m-1 \le \frac{N}{\sqrt{2}},$$

這裏取了正平方根；由於 $m \ge 1$，m 和 $m-1$ 都是非負數，所以這一步是成立的。因此 m 必須是滿足

$$\frac{N}{\sqrt{2}} \le m \le \frac{N}{\sqrt{2}} + 1.$$

的整數。

> ? **想一想** 長度為 1 的區間可以包含兩個整數。會不會有兩個中位數？

只有當區間的兩個端點都是整數時才會，而這需要 $N/\sqrt{2}$ 是整數。但若 $N/\sqrt{2} = k$，便有 $\sqrt{2} = N/k$，是一個有理數（rational number），這是不可能的。所以兩個不等式都是嚴格不等式，而長度為 1 的開區間（open interval）恰好包含一個整數，於是

$$m = \left\lceil \frac{N}{\sqrt{2}} \right\rceil,$$

即大於 $N/\sqrt{2}$ 的最小整數。

當 $N = 100$ 時，$N/\sqrt{2} = 50\sqrt{2} \approx 70.71$，所以 $m = 71$。直接驗證：$71^2 = 5041 \ge 5000 = \frac{1}{2}\times 100^2$，而 $70^2 = 4900 \le 5000$，所以兩個條件都成立。

第 (iv) 部分

由第 (iii) 部分，$m = N/\sqrt{2} + \theta$，其中 $0 < \theta < 1$，所以

$$\frac{m}{N} = \frac{1}{\sqrt{2}} + \frac{\theta}{N} \to \frac{1}{\sqrt{2}} \quad \text{當} N \to \infty.$$

由第 (ii) 部分，

$$\frac{\mu}{N} = \frac{(N+1)(4N-1)}{6N^2} = \frac{1}{6}\left(1+\frac{1}{N}\right)\left(4-\frac{1}{N}\right) \to \frac{4}{6} = \frac{2}{3}.$$

因此，當 N 很大時，

$$\frac{\mu}{m} = \frac{\mu/N}{m/N} \approx \frac{2/3}{1/\sqrt{2}} = \frac{2\sqrt{2}}{3}.$$

驗算

- 取 $N = 2$：四對等可能的結果中，$X = 1$ 出現一次，$X = 2$ 出現三次，與第 (i) 部分的 $\frac{1}{4}$ 和 $\frac{3}{4}$ 相符；而 $\mu = \frac{7}{4}$，與 $\frac{3\times 7}{12}$ 相符。
- 對 $N \le 40$，用電腦列出全部 N^2 對結果，可確認第 (i) 和第 (ii) 部分的結果完全正確；對 $N \le 2000$，把每個整數逐一代入中位數的定義檢驗，每次都恰好找到一個中位數，並且等於 $\lceil N/\sqrt{2} \rceil$。
- 當 $N = 10^5$ 時，$\mu/m \approx 0.942812$，而 $2\sqrt{2}/3 \approx 0.942809$。

常見錯誤

- 把 $\mathrm{P}(X = r)$ 算成 $2 \times \frac{r}{N} \cdot \frac{1}{N}$，忘記了這樣會把 $X_1 = X_2 = r$ 的情況數了兩次；正確的數目是 $2(r-1)+1 = 2r-1$。
- 寫成 $\mathrm{P}(X \ge m) = 1 - m^2/N^2$：這其實是 $\mathrm{P}(X > m)$，而不是 $\mathrm{P}(X \ge m)$。
- 在 $N = 100$ 時答 $m = 70$，原因是把 70.71 四捨五入至最接近的整數，而不是向上取整。

STEP 之外

X 是*順序統計量*（order statistic）最簡單的例子：樣本的最大值。當 $N \to \infty$ 時，X/N 趨近一個在 $[0,1]$ 上的連續隨機變量（continuous random variable），其密度函數為 $2x$，平均值是 $\frac{2}{3}$，中位數是 $1/\sqrt{2}$，這正是第 (iv) 部分出現該比值的原因。較大樣本的順序統計量會引出貝塔分佈（beta distribution）。Blitzstein 與 Hwang 合著的 *Introduction to Probability*（概率論入門）以淺白易明的方式講解順序統計量。

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。