詳細解答

STEP III 2022，第 11 題

概率與統計 • 二項分佈的平均偏差

STEP III 2022，第 11 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 OCR 提供下載的 2022 年 STEP 試卷（zip 檔案）內的 STEP III 試卷：

www.ocr.org.uk/Images/696587-2022-past-papers.zip

關鍵思路

硬幣公平，所以 X 的分佈（distribution）關於其均值（mean）對稱（symmetric），即低於均值的概率與高於均值的概率一一對應。利用這種對稱性，可以把絕對值（absolute value）化為單側的和。接着，(i)(c) 中的恆等式（identity）把 $\sum r\binom{2n}{r}$ 化為一個較小的二項分佈（binomial distribution）的尾部和（tail sum），而這個尾部和同樣可以用對稱性求出。

解答

全題中 $X \sim \mathrm{B}(N, \frac{1}{2})$，所以 $\mathrm{P}(X = r) = \binom{N}{r}2^{-N}$，且 $\mu = \frac{1}{2}N$。由於 $\binom{N}{r} = \binom{N}{N-r}$，我們有 $\mathrm{P}(X = r) = \mathrm{P}(X = N - r)$：分佈關於 $\frac{1}{2}N$ 對稱。

第 (i)(a) 部分

當 $N = 2n$ 時，映射 $r \mapsto 2n - r$ 把 $\{0, \ldots, n-1\}$ 映到 $\{n+1, \ldots, 2n\}$，所以由對稱性得

$$\mathrm{P}(X \leq n-1) = \mathrm{P}(X \geq n+1).$$

三個事件 $X \leq n-1$、$X = n$、$X \geq n+1$ 涵蓋了所有結果，所以 $2\,\mathrm{P}(X \leq n-1) + \mathrm{P}(X = n) = 1$，由此得

$$\mathrm{P}(X \leq n-1) = \tfrac{1}{2}(1 - \mathrm{P}(X = n)).$$

第 (i)(b) 部分

此時 $\mu = n$，所以

$$\delta = \sum_{r=0}^{2n} |r-n| \binom{2n}{r} \frac{1}{2^{2n}}.$$

$r = n$ 的項為零。在 $r \geq n+1$ 的各項中代入 $r = 2n - s$，得 $|r-n| = n - s$ 及 $\binom{2n}{r} = \binom{2n}{s}$，其中 $0 \leq s \leq n-1$。因此，高於均值的各項恰好重複了低於均值的各項：

$$\delta = 2\sum_{r=0}^{n-1} (n-r)\binom{2n}{r}\frac{1}{2^{2n}} = \sum_{r=0}^{n-1}(n-r)\binom{2n}{r}\frac{1}{2^{2n-1}},$$

得證。

第 (i)(c) 部分

對於 $1 \leq r \leq 2n$，

$$r\binom{2n}{r} = \frac{r\,(2n)!}{r!\,(2n-r)!} = 2n \cdot \frac{(2n-1)!}{(r-1)!\,(2n-r)!} = 2n\binom{2n-1}{r-1}.$$

現在把 (b) 中的和拆成 $2^{2n-1}\delta = A - B$，其中

$$A = n\sum_{r=0}^{n-1}\binom{2n}{r}, \qquad B = \sum_{r=1}^{n-1} r\binom{2n}{r}$$

（B 中 $r = 0$ 的項為零，所以略去）。

和 A。由 (a)，$\sum_{r=0}^{n-1}\binom{2n}{r} = 2^{2n}\,\mathrm{P}(X \leq n-1) = 2^{2n-1} - \frac{1}{2}\binom{2n}{n}$，所以

$$A = n\,2^{2n-1} - \tfrac{n}{2}\binom{2n}{n}.$$

想一想 當 s 由 0 到 $n-2$ 時，怎樣求 $\binom{2n-1}{s}$ 的總和？

和 B。由上述恆等式，令 $s = r - 1$，

$$B = 2n\sum_{s=0}^{n-2}\binom{2n-1}{s}.$$

這一行 $\binom{2n-1}{s}$（$0 \leq s \leq 2n-1$）共有偶數個項，總和為 2^{2n-1}。它在 $s \mapsto 2n-1-s$ 下對稱，而這個映射把 $\{0,\ldots,n-1\}$ 與 $\{n,\ldots,2n-1\}$ 互換，所以前一半的和是 2^{2n-2}。我們的和比這一半少了最後一項，所以

$$\sum_{s=0}^{n-2}\binom{2n-1}{s} = 2^{2n-2} - \binom{2n-1}{n-1} = 2^{2n-2} - \tfrac{1}{2}\binom{2n}{n},$$

這裏用了 $\binom{2n}{n} = \binom{2n-1}{n-1} + \binom{2n-1}{n} = 2\binom{2n-1}{n-1}$。因此

$$B = n\,2^{2n-1} - n\binom{2n}{n}.$$

合併起來。

$$2^{2n-1}\delta = A - B = \left(n - \frac{n}{2}\right)\binom{2n}{n} = \frac{n}{2}\binom{2n}{n}, \qquad \text{所以} \qquad \delta = \frac{n}{2^{2n}}\binom{2n}{n}.$$

第 (ii) 部分

現在 $N = 2n+1$，$\mu = n + \frac{1}{2}$，這並不是 X 可取的值。對稱映射 $r \mapsto 2n+1-r$ 把 $\{0,\ldots,n\}$ 與 $\{n+1,\ldots,2n+1\}$ 互換，並且保持 $|r-\mu|$ 和 $\binom{2n+1}{r}$ 不變。所以，與 (i)(b) 一樣，

$$\delta = 2\sum_{r=0}^{n}\left(n+\tfrac{1}{2}-r\right)\binom{2n+1}{r}\frac{1}{2^{2n+1}} = \frac{1}{2^{2n+1}}\sum_{r=0}^{n}(2n+1-2r)\binom{2n+1}{r}.$$

- 由同樣的對稱性，這一行 $\binom{2n+1}{r}$ 前一半的和是 2^{2n+1} 的一半：$\sum_{r=0}^{n}\binom{2n+1}{r} = 2^{2n}$。
- 由恆等式 $r\binom{2n+1}{r} = (2n+1)\binom{2n}{r-1}$（證明與 (i)(c) 完全相同），再用 (i)(a)，

$$\sum_{r=1}^{n} r\binom{2n+1}{r} = (2n+1)\sum_{s=0}^{n-1}\binom{2n}{s} = (2n+1)\left(2^{2n-1} - \tfrac{1}{2}\binom{2n}{n}\right).$$

因此

$$\sum_{r=0}^{n}(2n+1-2r)\binom{2n+1}{r} = (2n+1)2^{2n} - (2n+1)\left(2^{2n} - \binom{2n}{n}\right) = (2n+1)\binom{2n}{n},$$

而當 $N = 2n+1$ 時，

$$\boxed{\delta = \frac{2n+1}{2^{2n+1}}\binom{2n}{n}}$$

驗算

- $N = 3$（$n = 1$）：$|X - \frac{3}{2}|$ 以概率 $\frac{1}{4}$ 取值 $\frac{3}{2}$，以概率 $\frac{3}{4}$ 取值 $\frac{1}{2}$，所以 $\delta = \frac{3}{8} + \frac{3}{8} = \frac{3}{4}$。公式給出 $\frac{3\cdot 2}{8} = \frac{3}{4}$。當 $n = 0$（$N = 1$）時，公式給出 $\frac{1}{2}$，同樣正確。
- $N = 4$（$n = 2$）：(i)(c) 給出 $\frac{2\cdot 6}{16} = \frac{3}{4}$，與 $N = 3$ 的結果相同。事實上，$N = 2n+1$ 時的 δ 總是等於 $N = 2n+2$ 時的 δ：由於 $\binom{2n+2}{n+1} = \frac{2(2n+1)}{n+1}\binom{2n}{n}$，把偶數情況公式中的 n 換成 $n+1$，便化簡為奇數情況的公式。
- 經電腦驗算，對所有 $N \le 60$，兩條公式都與 δ 的定義完全吻合；例如 $N = 10$ 給出 $\delta = \frac{315}{256} \approx 1.230$。

常見錯誤

- 忘記對稱性帶來的因子 2；(i)(b) 中出現 2^{2n-1} 而不是 2^{2n}，正是因為這個因子。
- 作平移 $s = r - 1$ 後弄錯 s 的範圍：和止於 $n - 2$，比半行少一項，而正是缺少的那一項 $\binom{2n-1}{n-1}$ 令答案不為零。
- 在 (ii) 中把 μ 當作整數：其實沒有任何一項為零，而且這一行的兩半各有 $n + 1$ 項。

STEP 之外

斯特林公式（Stirling's formula）給出 $\binom{2n}{n} \approx 4^n/\sqrt{\pi n}$，所以當 N 很大時，$\delta \approx \sqrt{N/(2\pi)}$。由於 $\sigma = \frac{1}{2}\sqrt{N}$，這就是 $\delta \approx \sigma\sqrt{2/\pi}$，恰好是正態分佈（normal distribution）的平均絕對偏差（mean absolute deviation），正如中心極限定理（central limit theorem）所預測。Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率導論》）既講解了二項分佈的正態近似，也講解了其背後的中心極限定理。

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。