詳細解答

STEP III 2016，第 12 題

概率與統計 • 切比雪夫不等式

> **STEP III 2016，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/16-s3-q12

關鍵思路

切比雪夫不等式（Chebyshev's inequality）給出的是遠離均值（mean）的概率的上界，所以要證明某個量接近均值，就要從補事件（complement）入手。每一部分的工作都是：找出合適的隨機變量（random variable），求出它的均值和標準差（standard deviation），再選取 k，使 $k\sigma$ 恰好等於我們關心的距離。在第 (ii) 部分，和式 $\sum n^r/r!$ 與 e^n 並列出現，正是泊松分佈（Poisson distribution）的標誌。

解答

第 (i) 部分

設 X 為 $100n$ 次拋擲中得到正面的次數。假設各次拋擲互相獨立（independent），而每次得到正面的概率為 0.2，則 $X \sim \mathrm{B}(100n, 0.2)$，因此

$$\mu = 100n \times 0.2 = 20n, \qquad \sigma^2 = 100n \times 0.2 \times 0.8 = 16n, \qquad \sigma = 4\sqrt{n}.$$

事件 $16n \le X \le 24n$ 正好就是 $|X - 20n| \le 4n$，所以

$$\alpha = \mathrm{P}(|X - \mu| \le 4n) = 1 - \mathrm{P}(|X - \mu| > 4n).$$

我們希望 $k\sigma = 4n$，即 $k \cdot 4\sqrt{n} = 4n$，故 $k = \sqrt{n}$，這是正數。由切比雪夫不等式得

$$\mathrm{P}(|X - \mu| > 4n) = \mathrm{P}(|X - \mu| > \sqrt{n}\,\sigma) \le \frac{1}{(\sqrt{n})^2} = \frac{1}{n},$$

因此

$$\alpha \ge 1 - \frac{1}{n},$$

得證。（條件 $n > 1$ 只是為了讓這個界有實質內容：當 $n = 1$ 時，它只說 $\alpha \ge 0$。）

第 (ii) 部分

> **想一想** 哪一個隨機變量的概率形如 $n^r/r!$？

若 $Y \sim \mathrm{Po}(n)$，則對 $r = 0, 1, 2, \ldots$ 有 $\mathrm{P}(Y = r) = \mathrm{e}^{-n} n^r / r!$，所以

$$\mathrm{e}^{-n}\left(1 + n + \frac{n^2}{2!} + \cdots + \frac{n^{2n}}{(2n)!}\right) = \mathrm{P}(Y \le 2n).$$

所求不等式兩邊除以 $\mathrm{e}^n > 0$ 後，便等價於

$$\mathrm{P}(Y \le 2n) \ge 1 - \frac{1}{n}.$$

泊松分佈的均值和方差（variance）都是 n，所以 $\mu = n$，$\sigma = \sqrt{n}$。

> **想一想** 切比雪夫不等式是雙側的，但 $Y \le 2n$ 是單側事件。兩者如何對應起來？

由於 $Y \ge 0$，必有 $Y - n \ge -n$，所以 $Y - n < -n$ 不可能發生。因此

$$|Y - n| > n \iff Y - n > n \iff Y > 2n.$$

取 $k = \sqrt{n}$，使 $k\sigma = n$，由切比雪夫不等式得

$$\mathrm{P}(Y > 2n) = \mathrm{P}(|Y - n| > \sqrt{n}\,\sigma) \le \frac{1}{n},$$

所以 $\mathrm{P}(Y \le 2n) = 1 - \mathrm{P}(Y > 2n) \ge 1 - \frac{1}{n}$。兩邊乘以 e^n，

$$1 + n + \frac{n^2}{2!} + \cdots + \frac{n^{2n}}{(2n)!} \ge \left(1 - \frac{1}{n}\right)\mathrm{e}^n,$$

這正是所求結果。

驗算

這些界是正確的，但相當寬鬆。當 $n = 2$ 時，第 (i) 部分的準確二項分佈（binomial distribution）概率為 $\alpha \approx 0.868$，而界只是 0.5；當 $n = 10$ 時，$\alpha \approx 0.9986$，而界是 0.9。在第 (ii) 部分取 $n = 2$，左邊為 $1 + 2 + 2 + \frac{4}{3} + \frac{2}{3} = 7$，而 $\frac{1}{2}\mathrm{e}^2 \approx 3.69$。我們已就第 (i) 部分中 n 由 2 至 30、第 (ii) 部分中 n 由 1 至 59 的每一個值，準確驗證了兩個不等式。

常見錯誤

- 選取 k 時，誤用方差 $16n$ 代替標準差 $4\sqrt{n}$。

- 忘記切比雪夫不等式給出的是落在區間以外的概率的上界，因此要用補事件才能得到 α。
- 在第 (ii) 部分沒有解釋為何下尾 $Y < 0$ 沒有任何貢獻。

STEP 之外

切比雪夫不等式是集中不等式（concentration inequality）最簡單的例子；把它應用於 m 次獨立試驗的平均值，便得到弱大數定律（weak law of large numbers）：離散程度以 $1/\sqrt{m}$ 的速度收窄。第 (ii) 部分還告訴我們，指數級數（exponential series）在 $n^{2n}/(2n)!$ 之後的尾項之和至多為 e^n/n——這是一個用概率證明的分析學事實。更精細的工具，例如切爾諾夫界（Chernoff bounds），可以證明真正的概率與 1 的差距是指數級地小。如欲深入了解，可參閱 Grimmett 與 Welsh 的 *Probability: An Introduction*（《概率論導引》）。

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。