詳細解答

STEP II 2022，第 12 題

概率與統計 • 平均值、中位數與眾數

STEP II 2022，第 12 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 II 試卷：

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

關鍵思路

一切都源自累積分佈函數（cumulative distribution function）F，而這裏它只是一個簡單的多項式。由於 F 在 $[0,1]$ 上遞增，數 a 位於中位數（median）之下，當且僅當 $\mathrm{F}(a) < \frac{1}{2}$。因此，要比較平均值（mean）或眾數（mode）與中位數的大小，我們根本不需要求出中位數本身：只需計算 F 在平均值或眾數處的值，再與 $\frac{1}{2}$ 比較即可。

解答

第 (i) 部分

由於總概率為 1，

$$1 = k\int_0^1 \left(x^n - x^{n+1}\right)\mathrm{d}x = k\left(\frac{1}{n+1} - \frac{1}{n+2}\right) = \frac{k}{(n+1)(n+2)},$$

所以 $k = (n+1)(n+2)$，正如所求。於是

$$\mu = k\int_0^1 \left(x^{n+1} - x^{n+2}\right)\mathrm{d}x = (n+1)(n+2)\cdot\frac{1}{(n+2)(n+3)} = \frac{n+1}{n+3}.$$

第 (ii) 部分

當 $0 \le x \le 1$ 時，

$$\mathrm{F}(x) = (n+1)(n+2)\left(\frac{x^{n+1}}{n+1} - \frac{x^{n+2}}{n+2}\right) = x^{n+1}((n+2) - (n+1)x).$$

由於在 $(0,1)$ 上 $\mathrm{f} > 0$，F 在該區間上嚴格遞增（strictly increasing），而中位數 m 是唯一滿足 $\mathrm{F}(m) = \frac{1}{2}$ 的點。因此

$$\mu < m \iff \mathrm{F}(\mu) < \tfrac{1}{2}.$$

現在代入 $\mu = \frac{n+1}{n+3}$，

$$(n+2) - (n+1)\mu = \frac{(n+2)(n+3) - (n+1)^2}{n+3} = \frac{3n+5}{n+3},$$

所以

$$\mathrm{F}(\mu) = \left(\frac{n+1}{n+3}\right)^{n+1} \frac{3n+5}{n+3}.$$

因此，條件 $\mathrm{F}(\mu) < \frac{1}{2}$ 兩邊乘以正數 $2\left(\frac{n+3}{n+1}\right)^{n+1}$ 後，變為

$$\frac{2(3n+5)}{n+3} < \left(\frac{n+3}{n+1}\right)^{n+1}.$$

由於 $\frac{2(3n+5)}{n+3} = \frac{6(n+3)-8}{n+3} = 6 - \frac{8}{n+3}$ 而且 $\frac{n+3}{n+1} = 1 + \frac{2}{n+1}$，這正是題目給出的不等式。所以只要該不等式成立，$\mu$ 就小於中位數。

要證明這個不等式，設 $N = n+1$，於是 $N \geq 3$ 是整數。$\left(1 + \frac{2}{N}\right)^N$ 的二項展開（binomial expansion）中每一項都是正數，而且共有 $N+1 \geq 4$ 項，所以總和至少等於首四項之和：

$$\begin{aligned}\left(1+\frac{2}{N}\right)^N &\geq 1 + N\cdot\frac{2}{N} + \frac{N(N-1)}{2}\cdot\frac{4}{N^2} + \frac{N(N-1)(N-2)}{6}\cdot\frac{8}{N^3} \\ &= 3 + 2\left(1 - \frac{1}{N}\right) + \frac{4}{3}\left(1 - \frac{3}{N} + \frac{2}{N^2}\right) \\ &= \frac{19}{3} - \frac{6}{N} + \frac{8}{3N^2}.\end{aligned}$$

想一想 這個四項之和只是一個下界。我們怎樣證明它大於 $6 - \frac{8}{N+2}$？

我們把兩者相減，再乘以正的分母 $3N^2(N+2)$ 以消去分母：

$$\begin{aligned}&3N^2(N+2)\left[\frac{19}{3} - \frac{6}{N} + \frac{8}{3N^2} - 6 + \frac{8}{N+2}\right] \\ &\quad = N^2(N+2) - 18N(N+2) + 8(N+2) + 24N^2 \\ &\quad = N^3 + 8N^2 - 28N + 16.\end{aligned}$$

當 $N \geq 3$ 時，我們有 $N^2 + 8N - 28 \geq 9 + 24 - 28 = 5 > 0$，所以 $N^3 + 8N^2 - 28N + 16 = N(N^2 + 8N - 28) + 16 > 0$。因此

$$6 - \frac{8}{n+3} < \frac{19}{3} - \frac{6}{N} + \frac{8}{3N^2} \leq \left(1 + \frac{2}{n+1}\right)^{n+1},$$

即題目給出的不等式對每個整數 $n > 1$ 都成立，所以 X 的中位數大於 μ。

第 (iii) 部分

在 $(0,1)$ 上，

$$\mathrm{f}'(x) = k\left(nx^{n-1} - (n+1)x^n\right) = kx^{n-1}(n-(n+1)x),$$

當 $x < \frac{n}{n+1}$ 時它是正數，當 $x > \frac{n}{n+1}$ 時它是負數。所以 f 先升後降，眾數為 $M = \frac{n}{n+1}$。與第 (ii) 部分一樣，$M > m$ 當且僅當 $\mathrm{F}(M) > \frac{1}{2}$。由於 $(n+2) - (n+1)M = (n+2) - n = 2$，

$$\mathrm{F}(M) = 2\left(\frac{n}{n+1}\right)^{n+1},$$

所以我們需要 $\left(\frac{n}{n+1}\right)^{n+1} > \frac{1}{4}$，即

$$\left(1+\frac{1}{n}\right)^{n+1} < 4.$$

? **想一想** 既然 n 可以是 2 或以上的任何整數，題目給出關於 $\left(1+\frac{1}{x}\right)^{x+1}$ 的事實怎樣幫到我們？

由於該函數在 x 為正數時遞減，它在任何整數 $n \geq 2$ 處的值都不大於它在 $x = 2$ 處的值：

$$\left(1+\frac{1}{n}\right)^{n+1} \leq \left(\frac{3}{2}\right)^3 = \frac{27}{8} < 4.$$

因此 $\mathrm{F}(M) > \frac{1}{2}$，即 X 的眾數大於其中位數。

綜合各部分，對每個 $n > 1$ 都有 平均值 $<$ 中位數 $<$ 眾數。

驗算

- 取 $n = 2$：平均值為 $\frac{3}{5} = 0.6$，眾數為 $\frac{2}{3} \approx 0.667$。以數值方法解 $4x^3 - 3x^4 = \frac{1}{2}$，得中位數 ≈ 0.6143，恰好介乎兩者之間。此外，$\mathrm{F}(0.6) = 0.4752 < \frac{1}{2}$，而 $\mathrm{F}(\frac{2}{3}) = \frac{16}{27} \approx 0.593 > \frac{1}{2}$。
- 當 $n = 2\,(N = 3)$ 時，展開式剛好有四項，所以下界取等號：兩邊都等於 $\frac{125}{27} \approx 4.63$，穩穩大於 $6 - \frac{8}{5} = 4.4$。
- 平均值 $<$ 中位數 $<$ 眾數這個大小次序，已用電腦對 2 至 79 的每個 n 驗證過。

常見錯誤

- 試圖明確求出中位數：方程 $x^{n+1}(n+2-(n+1)x) = \frac{1}{2}$ 沒有簡潔的解，而你也不需要求出它。

- 弄錯不等號方向：由於 F 遞增，$\mu < m$ 對應的是 $\mathrm{F}(\mu) < \frac{1}{2}$。
- 忘記說明為何可以捨去後面的二項式項（因為它們全是正數）。

STEP 之外

這個密度函數是一個貝塔分佈（Beta distribution）Beta$(n+1, 2)$，它是 $[0,1]$ 上最重要的分佈族之一，在貝葉斯統計（Bayesian statistics）中處於核心地位。平均值 < 中位數 < 眾數這個次序，是左邊有長尾（負偏態，negative skew）的分佈的典型特徵，但它並非普遍定律，所以這裏必須加以證明。第 (iii) 部分中的遞減函數在 $x \to \infty$ 時趨向 e，這是一個經典極限，你將會在分析（analysis）課程中正式學到。Blitzstein 與 Hwang 的 *Introduction to Probability*（《概率導論》）對貝塔分佈有出色的論述。

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。