詳細解答

STEP II 2019，第 12 題

概率與統計 • 連續分佈與分位數

STEP II 2019，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 2019 年 STEP 試卷（zip 檔案）內的 STEP II 試卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

關鍵思路

這裏的累積分佈函數（cumulative distribution function）在 $[0,1]$ 上就是 x^n，所以每個分位數（quantile）都可以準確寫出：分佈中有比例 p 落在其下方的值是 $p^{1/n}$。於是，比較平均值與某個分位數就變成比較冪，而把兩邊同時取 n 次方後，問題便化為關於 $\left(1+\frac{1}{n}\right)^n$ 的命題，這正是二項式展開（binomial expansion）最擅長估計的量。

解答

第 (i) 部分

對概率密度函數（probability density function）積分，

$$\mu = \mathrm{E}(X) = \int_0^1 x \cdot nx^{n-1}\,\mathrm{d}x = \int_0^1 nx^n\,\mathrm{d}x = \frac{n}{n+1},$$
$$\mathrm{E}(X^2) = \int_0^1 nx^{n+1}\,\mathrm{d}x = \frac{n}{n+2}.$$

因此

$$\sigma^2 = \mathrm{E}(X^2) - \mu^2 = \frac{n}{n+2} - \frac{n^2}{(n+1)^2} = \frac{n\left[(n+1)^2 - n(n+2)\right]}{(n+1)^2(n+2)}$$
$$= \frac{n\left[n^2+2n+1-n^2-2n\right]}{(n+1)^2(n+2)} = \frac{n}{(n+1)^2(n+2)},$$

得證。

第 (ii) 部分

當 $0 \le x \le 1$ 時，分佈函數為

$$\mathrm{F}(x) = \mathrm{P}(X \le x) = \int_0^x nt^{n-1}\,\mathrm{d}t = x^n.$$

取 $n = 2$，則 $\mathrm{F}(x) = x^2$。下四分位數（lower quartile）q_1 滿足 $q_1^2 = \frac{1}{4}$，上四分位數（upper quartile）q_3 滿足 $q_3^2 = \frac{3}{4}$；兩者都在 $[0,1]$ 內，所以取正根：

$$q_1 = \tfrac{1}{2}, \qquad q_3 = \tfrac{\sqrt{3}}{2}, \qquad q_3 - q_1 = \frac{\sqrt{3}-1}{2}.$$

由第 (i) 部分，$\sigma^2 = \dfrac{2}{9\cdot 4} = \dfrac{1}{18}$，所以 $\sigma = \dfrac{1}{3\sqrt{2}} = \dfrac{\sqrt{2}}{6}$，而 $2\sigma = \dfrac{\sqrt{2}}{3}$。

我們要證明 $\dfrac{\sqrt{3}-1}{2} < \dfrac{\sqrt{2}}{3}$。兩邊乘以 6，這等價於

$$3\sqrt{3} - 3 < 2\sqrt{2} \iff 3\sqrt{3} < 3 + 2\sqrt{2}.$$

最後這個不等式兩邊都是正數，所以它等價於兩邊平方後的不等式：

$$27 < 9 + 12\sqrt{2} + 8 = 17 + 12\sqrt{2} \iff 10 < 12\sqrt{2},$$

這是成立的，因為 $\sqrt{2} > 1$ 給出 $12\sqrt{2} > 12 > 10$。每一步都是等價的，所以沿着這條推理鏈倒推回去，可知四分位數間距（interquartile range）小於 2σ。

> **想一想** 為甚麼在這裏兩邊平方是安全的，但一般情況下卻不是？

只有當 a 和 b 都非負時，平方才能保持不等式 $a < b$（例如 $-3 < 2$，但 $9 > 4$）。正因如此，我們先把不等式整理成 $3\sqrt{3} < 3 + 2\sqrt{2}$，使兩邊明顯都是正數，而不是直接把仍含有減法的 $3\sqrt{3} - 3 < 2\sqrt{2}$ 平方。

第 (iii) 部分

由二項式定理，

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{2}\,x^2 + \cdots,$$

而第 $(k+1)$ 項是

$$\binom{n}{k} x^k = \frac{n(n-1)\cdots(n-k+1)}{k!}\,x^k.$$

中位數與下四分位數。 由於 $\mathrm{F}(x) = x^n$，中位數（median）m 滿足 $m^n = \frac{1}{2}$，下四分位數 q 滿足 $q^n = \frac{1}{4}$，所以

$$m = \left(\tfrac{1}{2}\right)^{1/n}, \qquad q = \left(\tfrac{1}{4}\right)^{1/n}.$$

μ、m、q 全都是正數，而 $t \mapsto t^n$ 在 $t > 0$ 上嚴格遞增（strictly increasing），所以我們可以透過比較它們的 n 次方來比較它們。另外

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

記 $A = \left(1+\frac{1}{n}\right)^n$。那麼

$$\mu < m \iff \mu^n < \tfrac{1}{2} \iff A > 2, \qquad \mu > q \iff \mu^n > \tfrac{1}{4} \iff A < 4.$$

μ 小於中位數。 在展開式中代入 $x = \frac{1}{n}$。每一項都非負，所以 A 至少等於首三項之和：

$$A \geq 1 + n \cdot \frac{1}{n} + \frac{n(n-1)}{2} \cdot \frac{1}{n^2} = 2 + \frac{n-1}{2n} > 2,$$

其中最後一個不等式是嚴格的，因為 $n > 1$。（當 $n \geq 2$ 時，展開式中確實有第三項。）因此 $\mu < m$。

μ 大於下四分位數。 取 $x = \frac{1}{n}$，第 $(k+1)$ 項是

$$\binom{n}{k}\frac{1}{n^k} = \frac{1}{k!} \cdot \frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n} \leq \frac{1}{k!},$$

因為那 k 個分數 $\frac{n-j}{n}$（$0 \leq j \leq k-1$）每個都在 0 與 1 之間。（當 $k = 0$ 時，該項是 $1 = \frac{1}{0!}$。）對 $k = 0, 1, \ldots, n$ 求和，

$$A \leq 1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} < 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots < 4,$$

中間的不等式成立，是因為無窮級數（infinite series）還包含更多正項；最後一個不等式則是題目容許我們假設的結果。因此 $A < 4$，所以 $\mu > q$。

綜合兩者，對每個整數 $n > 1$，下四分位數都小於 μ，而 μ 又小於中位數。

想一想 為甚麼比較 n 次方就能確定 μ 與 m 的大小關係？

對正數 a 和 b，$a < b$ 若且唯若 $a^n < b^n$，因為 t^n 在 $t > 0$ 上嚴格遞增。我們的三個數全是正數，所以取 n 次方不會損失任何資訊，而且這樣做能消去 m 和 q 中麻煩的 n 次方根。

驗算

► 當 $n = 2$：$\mu = \frac{2}{3}$，而 $\sigma^2 = \frac{2}{9 \cdot 4} = \frac{1}{18}$，與 $\mathrm{E}(X^2) - \mu^2 = \frac{1}{2} - \frac{4}{9}$ 吻合。

- 數值上，當 $n = 2$ 時，四分位數間距是 0.366，而 $2\sigma = 0.471$。
- 當 $n = 2$ 時，三個值分別是 $q = 0.5$、$\mu \approx 0.667$、$m \approx 0.707$；當 $n = 10$ 時，它們是 $q \approx 0.871$、$\mu \approx 0.909$、$m \approx 0.933$。對 2 至 2000 的每個 n，順序 $q < \mu < m$ 都已用電腦驗證。
- 當 $n \to \infty$ 時，$A = \left(1 + \frac{1}{n}\right)^n \to \mathrm{e} \approx 2.718$，穩穩地落在 2 與 4 之間。

常見錯誤

- 沒有先確認兩邊都是正數，就直接把 $3\sqrt{3} - 3 < 2\sqrt{2}$ 平方。
- 取倒數時弄錯不等號方向：$\mu^n < \frac{1}{2}$ 對應的是 $A > 2$，而不是 $A < 2$。
- 聲稱第 $(k+1)$ 項等於 $\frac{1}{k!}$，而不是至多為 $\frac{1}{k!}$。

STEP 之外

$\left(1 + \frac{1}{n}\right)^n$ 隨 n 遞增並趨向極限 e；而你逐項與 $\sum 1/k!$ 比較所證出的界，正是證明 $\mathrm{e} = \sum_{k \geq 0} 1/k!$ 的第一步，這是大學第一門數學分析（analysis）課程中的標準結果。這個分佈本身是 Beta$(n, 1)$ 分佈（Beta distribution）；它也是 n 個在 $[0, 1]$ 上獨立均勻分佈的隨機變量之最大值的分佈，因為 $\mathrm{P}(\max \leq x) = x^n$。想深入了解這兩條線索，可參閱 Blitzstein 與 Hwang 的 *Introduction to Probability*（《概率導論》），以及 Abbott 的 *Understanding Analysis*（《理解分析》）。

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。