詳細解答

STEP II 2011，第 13 題

概率與統計 • 偏度的量度

> **STEP II 2011，第 13 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/11-s2-q13

關鍵思路

偏度（skewness）描述的是分佈的「不對稱」程度。這裏的三種量度從三個不同角度把它找出來：透過三階中心矩（third central moment）、透過中位數（median）在兩個十分位數（decile）之間的位置，以及透過平均值與中位數之間的差距。對於 $[0,1]$ 上的密度 $2x$，一切都可以明確算出：各階矩都只是一行就能完成的積分，而累積分佈函數（cumulative distribution function）是 x^2，所以它的反函數就是平方根。真正要花功夫的只是小心處理根式（surd）的運算，以及在最後找一個乾淨利落的方法比較它們的大小。

解答

偏度量度甚麼。偏度量度的是分佈相對其中心的*不對稱性*。對稱分佈的偏度為零。若較長的尾部在右邊，偏度為正；若較長的尾部在左邊，偏度為負。

第 (i) 部分

把立方展開，並利用期望值的線性性質（linearity of expectation），記住 μ 是常數：

$$\begin{aligned}\mathrm{E}((X-\mu)^3) &= \mathrm{E}(X^3) - 3\mu\,\mathrm{E}(X^2) + 3\mu^2\,\mathrm{E}(X) - \mu^3 \\ &= \mathrm{E}(X^3) - 3\mu(\sigma^2+\mu^2) + 3\mu^3 - \mu^3 \\ &= \mathrm{E}(X^3) - 3\mu\sigma^2 - \mu^3,\end{aligned}$$

這裏用了 $\mathrm{E}(X)=\mu$ 和 $\mathrm{E}(X^2)=\sigma^2+\mu^2$（即方差（variance）的常用公式移項後的形式）。除以 σ^3，便得到所求的 γ 表達式。

密度為 $2x$ 的分佈。對於 $k \geq 0$，

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

所以 $\mu = \frac{2}{3}$、$\mathrm{E}(X^2) = \frac{1}{2}$ 及 $\mathrm{E}(X^3) = \frac{2}{5}$。因此

$$\sigma^2 = \tfrac{1}{2} - \tfrac{4}{9} = \tfrac{1}{18}, \qquad \sigma = \frac{1}{3\sqrt{2}}, \qquad \sigma^3 = \frac{1}{27 \cdot 2\sqrt{2}} = \frac{1}{54\sqrt{2}}.$$

分子為

$$\mathrm{E}(X^3) - 3\mu\sigma^2 - \mu^3 = \frac{2}{5} - 3 \cdot \frac{2}{3} \cdot \frac{1}{18} - \frac{8}{27} = \frac{2}{5} - \frac{1}{9} - \frac{8}{27} = \frac{54 - 15 - 40}{135} = -\frac{1}{135}.$$

因此

$$\gamma = -\frac{1}{135} \cdot 54\sqrt{2} = -\frac{54\sqrt{2}}{135} = -\frac{2\sqrt{2}}{5},$$

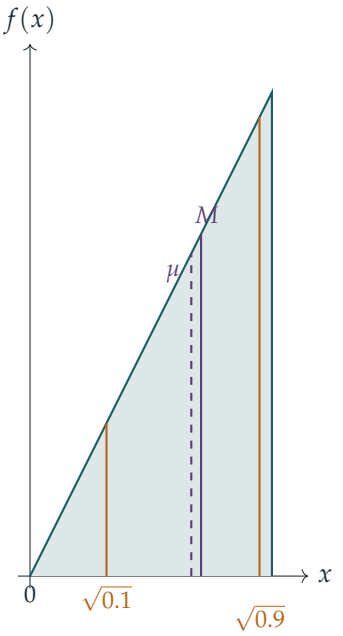
正如所求。負號是合理的：密度集中在 $x = 1$ 附近，在左邊留下一條伸向 0 的長尾。

第 (ii) 部分

對於 $0 \leq x \leq 1$，累積分佈函數為

$$\mathrm{F}(x) = \int_0^x 2t \,\mathrm{d}t = x^2,$$

它在 $[0,1]$ 上遞增，所以對於 $0 \leq p \leq 1$，其反函數為 $\mathrm{F}^{-1}(p) = \sqrt{p}$。



十分位數偏度。把 D 的分子和分母同乘以 $\sqrt{10}$，便可消去所有根式，只剩下一個：

$$D = \frac{\sqrt{\frac{9}{10}} - 2\sqrt{\frac{1}{2}} + \sqrt{\frac{1}{10}}}{\sqrt{\frac{9}{10}} - \sqrt{\frac{1}{10}}} = \frac{3 - 2\sqrt{5} + 1}{3 - 1} = \frac{4 - 2\sqrt{5}}{2} = 2 - \sqrt{5},$$

因為 $\sqrt{10} \cdot \sqrt{\frac{1}{2}} = \sqrt{5}$。這正是要證明的結果。

皮爾遜偏度。中位數滿足 $\mathrm{F}(M) = \frac{1}{2}$，所以 $M = \frac{1}{\sqrt{2}}$。利用 $\frac{1}{\sigma} = 3\sqrt{2}$，

$$P = 3\Big(\frac{2}{3} - \frac{1}{\sqrt{2}}\Big) \cdot 3\sqrt{2} = 9\sqrt{2} \cdot \frac{2}{3} - 9 = 6\sqrt{2} - 9.$$

數值上，$D \approx -0.236$、$P \approx -0.515$ 及 $\gamma \approx -0.566$，所以大小次序看來是對的；現在我們加以證明。

> **想一想** 不用計算機，怎樣比較 $2 - \sqrt{5}$ 和 $6\sqrt{2} - 9$ 這類數的大小？

先移項使兩邊都是正數，然後平方：對於正數，$a < b$ 當且僅當 $a^2 < b^2$。以下每一步都是可逆的，所以最後的陳述成立，便證明了最初的陳述。

$D > P$。我們需要證明 $2 - \sqrt{5} > 6\sqrt{2} - 9$，即 $\sqrt{5} + 6\sqrt{2} < 11$。兩邊都是正數，所以這等價於

$$\begin{aligned}(\sqrt{5} + 6\sqrt{2})^2 = 77 + 12\sqrt{10} < 121 &\iff 12\sqrt{10} < 44 \\ &\iff 3\sqrt{10} < 11 \\ &\iff 90 < 121,\end{aligned}$$

其中最後一步再次把兩個正數平方。由於 $90 < 121$ 成立，所以 $D > P$。

$P > \gamma$。我們需要證明 $6\sqrt{2} - 9 > -\frac{2\sqrt{2}}{5}$。乘以 5 並移項，即 $32\sqrt{2} > 45$。兩邊都是正數，所以這等價於 $32^2 \cdot 2 > 45^2$，即 $2048 > 2025$，這是成立的。所以 $P > \gamma$。

因此 $D > P > \gamma$。三種量度都是負數，一致認為這個分佈向左偏，但對偏的程度卻各有不同的看法。

驗算

- (i) 中的恆等式對任何對稱分佈都給出 0：例如，若 X 以相等概率取 ± 1，則 $\mu = 0$、$\sigma = 1$ 及 $\mathrm{E}(X^3) = 0$。
- 數值上，用中點法則計算 $\mathrm{E}\big((X-\mu)^3\big)/\sigma^3$，得到 -0.56569，與 $-2\sqrt{2}/5$ 吻合。
- 以 U 為 $[0,1]$ 上的均勻分佈，模擬 $X = \sqrt{U}$（它滿足 $\mathrm{F}(x) = x^2$），得到的樣本中位數接近 0.707，十分位數接近 0.316 和 0.949，與草圖一致。

- 平均值 $\frac{2}{3}$ 位於中位數 $\frac{1}{\sqrt{2}} \approx 0.707$ 的左邊，這是左偏分佈的典型特徵，也是 $P < 0$ 的原因。

常見錯誤

- 把 $\sigma^3 = (\frac{1}{18})^3$ 寫成這樣，而不是 $(\frac{1}{18})^{3/2}$：要立方的是標準差（standard deviation），不是方差。
- 在其中一邊可能是負數的情況下把不等式平方。應先移項使兩邊都是正數。
- 在 γ 的公式或 P 的分母中，該用 $\sigma^2 = \frac{1}{18}$ 的地方卻用了 $\mathrm{E}(X^2) = \frac{1}{2}$。

STEP 之外

$\mathrm{E}\big((X-\mu)^3\big)$ 這個量是三階中心矩，而 (i) 中的展開正是矩與累積量（cumulant）之間關係的起點；對於獨立隨機變量之和，累積量可以簡單地相加。像 D 這類以分位數（quantile）為基礎的量度在統計學中備受重視，因為它們是穩健的（robust）：單一個極端的觀測值幾乎不會移動十分位數，卻可以大幅改變三階矩。Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率導論》）以合適的程度講解矩與矩母函數（moment generating function）。

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。