详细解答

STEP II 2011，第 13 题

概率与统计 · 偏度的量度

> **STEP II 2011，第 13 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/11-s2-q13

关键思路

偏度（skewness）描述的是分布的“不对称”程度。这里的三种量度从三个不同角度把它找出来：通过三阶中心矩（third central moment）、通过中位数（median）在两个十分位数（decile）之间的位置，以及通过平均值与中位数之间的差距。对于 $[0,1]$ 上的密度 $2x$，一切都可以明确算出：各阶矩都只是一行就能完成的积分，而累积分布函数（cumulative distribution function）是 x^2，所以它的反函数就是平方根。真正要花功夫的只是小心处理根式（surd）的运算，以及在最后找一个干净利落的方法比较它们的大小。

解答

偏度量度甚么。偏度量度的是分布相对其中心的*不对称性*。对称分布的偏度为零。若较长的尾部在右边，偏度为正；若较长的尾部在左边，偏度为负。

第 (i) 部分

把立方展开，并利用期望值的线性性质（linearity of expectation），记住 μ 是常数：

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

这里用了 $\mathrm{E}(X)=\mu$ 和 $\mathrm{E}(X^2)=\sigma^2+\mu^2$（即方差（variance）的常用公式移项后的形式）。除以 σ^3，便得到所求的 γ 表达式。

密度为 $2x$ 的分布。对于 $k \geq 0$,

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

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

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

分子为

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

因此

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

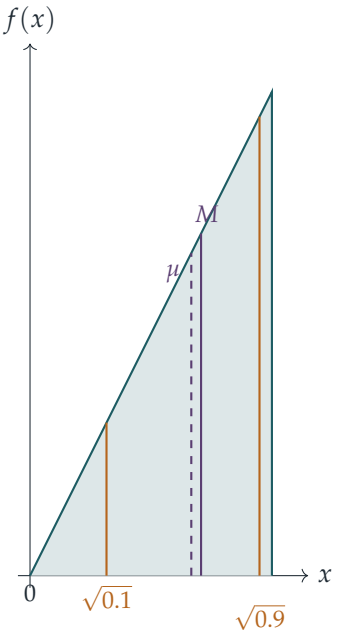
正如所求。负号是合理的：密度集中在 $x = 1$ 附近，在左边留下一条伸向 0 的长尾。

第 (ii) 部分

对于 $0 \leq x \leq 1$，累积分布函数为

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

它在 $[0,1]$ 上递增，所以对于 $0 \leq p \leq 1$，其反函数为 $\mathrm{F}^{-1}(p) = \sqrt{p}$。



十分位数偏度。把 D 的分子和分母同乘以 $\sqrt{10}$，便可消去所有根式，只剩下一个：

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

因为 $\sqrt{10} \cdot \sqrt{\frac{1}{2}} = \sqrt{5}$。这正是要证明的结果。

皮尔逊偏度。中位数满足 $\mathrm{F}(M) = \frac{1}{2}$，所以 $M = \frac{1}{\sqrt{2}}$。利用 $\frac{1}{\sigma} = 3\sqrt{2}$，

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

数值上，$D \approx -0.236$、$P \approx -0.515$ 及 $\gamma \approx -0.566$，所以大小次序看来是对的；现在我们加以证明。

> ? **想一想** 不用计算机，怎样比较 $2 - \sqrt{5}$ 和 $6\sqrt{2} - 9$ 这类数的大小?

先移项使两边都是正数，然后平方：对于正数，$a < b$ 当且仅当 $a^2 < b^2$。以下每一步都是可逆的，所以最后的陈述成立，便证明了最初的陈述。

$D > P$。我们需要证明 $2 - \sqrt{5} > 6\sqrt{2} - 9$，即 $\sqrt{5} + 6\sqrt{2} < 11$。两边都是正数，所以这等价于

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

其中最后一步再次把两个正数平方。由于 $90 < 121$ 成立，所以 $D > P$。

$P > \gamma$。我们需要证明 $6\sqrt{2} - 9 > -\frac{2\sqrt{2}}{5}$。乘以 5 并移项，即 $32\sqrt{2} > 45$。两边都是正数，所以这等价于 $32^2 \cdot 2 > 45^2$，即 $2048 > 2025$，这是成立的。所以 $P > \gamma$。

因此 $D > P > \gamma$。三种量度都是负数，一致认为这个分布向左偏，但对偏的程度却各有不同的看法。

验算

- (i) 中的恒等式对任何对称分布都给出 0：例如，若 X 以相等概率取 ± 1，则 $\mu = 0$、$\sigma = 1$ 及 $\mathrm{E}(X^3) = 0$。
- 数值上，用中点法则计算 $\mathrm{E}\big((X-\mu)^3\big)/\sigma^3$，得到 -0.56569，与 $-2\sqrt{2}/5$ 吻合。
- 以 U 为 $[0,1]$ 上的均匀分布，模拟 $X = \sqrt{U}$（它满足 $\mathrm{F}(x) = x^2$），得到的样本中位数接近 0.707，十分位数接近 0.316 和 0.949，与草图一致。

- 平均值 $\frac{2}{3}$ 位于中位数 $\frac{1}{\sqrt{2}} \approx 0.707$ 的左边，这是左偏分布的典型特征，也是 $P < 0$ 的原因。

常见错误

- 把 $\sigma^3 = (\frac{1}{18})^3$ 写成这样，而不是 $(\frac{1}{18})^{3/2}$：要立方的是标准差（standard deviation），不是方差。
- 在其中一边可能是负数的情况下把不等式平方。应先移项使两边都是正数。
- 在 γ 的公式或 P 的分母中，该用 $\sigma^2 = \frac{1}{18}$ 的地方却用了 $\mathrm{E}(X^2) = \frac{1}{2}$。

STEP 之外

$\mathrm{E}\big((X-\mu)^3\big)$ 这个量是三阶中心矩，而 (i) 中的展开正是矩与累积量（cumulant）之间关系的起点；对于独立随机变量之和，累积量可以简单地相加。像 D 这类以分位数（quantile）为基础的量度在统计学中备受重视，因为它们是稳健的（robust）：单一个极端的观测值几乎不会移动十分位数，却可以大幅改变三阶矩。Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率导论》）以合适的程度讲解矩与矩母函数（moment generating function）。

Beyond STEP 详细解答：https://beyondstep.keithmcnulty.org/zh-hans/。解答文字采用 CC0 1.0；题目本身版权属 Cambridge University Press & Assessment 所有（©），本文件不予转载。由机器从英文翻译而成；尚未经审阅。