详细解答

STEP III 2016，第 13 题

概率与统计 · 和的峰度

STEP III 2016，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/16-s3-q13

关键思路

峰度（kurtosis）只由偏离平均值的差构成，所以任何平移变量的操作都不会改变它；这让我们可以把每个变量都中心化至 0，然后直接处理普通的矩（moments）。对于正态分布（Normal distribution），四阶矩可用分部积分法（integration by parts）求得，方法是分出一个 x 因子与指数函数配对。对于若干个互相独立（independent）且平均值为零的变量之和，展开 T^4 再取期望值，凡是含有某个变量的一次方的项都会变成零。这就是第 (ii) 部分的全部内容，而它会直接用于第 (iii) 部分。

解答

平移。 设 $W = X - a$。由期望值的线性性质（linearity of expectation），$\mathrm{E}(W) = \mu - a$，所以

$$W - \mathrm{E}(W) = (X - a) - (\mu - a) = X - \mu.$$

W 偏离其平均值的差，正好等于 X 偏离其平均值的差。因此 W 有相同的方差 $\mathrm{E}((X-\mu)^2) = \sigma^2$、相同的四阶中心矩（central moment）$\mathrm{E}((X-\mu)^4)$，从而有相同的峰度。特别地，取 $a = \mu$，则每个随机变量都与其中心化版本 $X - \mu$ 有相同的峰度，而后者的平均值为 0。

第 (i) 部分

设 $X \sim \mathrm{N}(0, \sigma^2)$，其概率密度函数（probability density function）为

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\,\mathrm{e}^{-x^2/(2\sigma^2)}, \qquad \text{因此} \qquad f'(x) = -\frac{x}{\sigma^2}\,f(x).$$

由于 $\mu = 0$，我们需要求 $\mathrm{E}(X^4) = \int_{-\infty}^{\infty} x^4 f(x)\,\mathrm{d}x$。

想一想 既然 $\mathrm{e}^{-x^2/(2\sigma^2)}$ 没有初等的原函数，我们怎样积分 $x^4\mathrm{e}^{-x^2/(2\sigma^2)}$?

函数 $xf(x)$ 确实有原函数（antiderivative），就是 $-\sigma^2 f(x)$。所以写成 $x^4 f(x) = x^3 \cdot xf(x)$，并取 $u = x^3$ 及 $\mathrm{d}v = xf(x)\,\mathrm{d}x$、$v = -\sigma^2 f(x)$ 进行分部积分：

$$\int_{-\infty}^{\infty} x^4 f(x)\,\mathrm{d}x = \Big[-\sigma^2 x^3 f(x)\Big]_{-\infty}^{\infty} + 3\sigma^2 \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x = 3\sigma^2 \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x.$$

边界项为零，因为当 $x \to \pm\infty$ 时 $x^3\mathrm{e}^{-x^2/(2\sigma^2)} \to 0$（指数函数衰减的速度比任何幂的增长都快）。以 $u = x$ 重复同一步骤，得

$$\int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x = \Big[-\sigma^2 x f(x)\Big]_{-\infty}^{\infty} + \sigma^2 \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = \sigma^2,$$

因为 f 是概率密度函数，其积分为 1。（这也证实了方差是 σ^2。）因此 $\mathrm{E}(X^4) = 3\sigma^2 \cdot \sigma^2 = 3\sigma^4$，而

$$\kappa = \frac{3\sigma^4}{\sigma^4} - 3 = 0.$$

由平移的结果，每个正态分布，无论其平均值是多少，峰度都是 0。

第 (ii) 部分

用多项式定理（multinomial theorem）展开 $T^4 = (Y_1 + \cdots + Y_n)^4$。每一项都是四个因子 Y_r 的乘积，而按照这四个因子如何分配到不同的下标，各项可分为五类：

$$Y_r^4, \qquad 4\,Y_r^3 Y_s, \qquad 6\,Y_r^2 Y_s^2, \qquad 12\,Y_r^2 Y_s Y_t, \qquad 24\,Y_r Y_s Y_t Y_u,$$

其中不同的字母代表不同的下标。系数计算的是排列方式的数目：例如，$Y_r^2 Y_s^2$ 来自从 4 个括号中选出哪 2 个提供 Y_r，共有 $\binom{4}{2} = 6$ 种方法。每个无序对 $\{r, s\}$ 在 $Y_r^2 Y_s^2$ 这一类中只出现一次。

现在取期望值。各个 Y 互相独立，所以不同 Y 的幂的乘积的期望值，等于各自期望值的乘积。在 $Y_r^3 Y_s$、$Y_r^2 Y_s Y_t$ 和 $Y_r Y_s Y_t Y_u$ 这几类中，总有某个变量只以一次方出现，而它的因子 $\mathrm{E}(Y_s) = 0$ 令整个期望值为零。例如，

$$\mathrm{E}(Y_r^3 Y_s) = \mathrm{E}(Y_r^3)\,\mathrm{E}(Y_s) = 0.$$

只有第一类和第三类保留下来，而同样由独立性，当 $r \neq s$ 时 $\mathrm{E}(Y_r^2 Y_s^2) = \mathrm{E}(Y_r^2)\,\mathrm{E}(Y_s^2)$。把各对写成 $r < s$，得

$$\mathrm{E}(T^4) = \sum_{r=1}^{n} \mathrm{E}(Y_r^4) + 6 \sum_{r=1}^{n-1} \sum_{s=r+1}^{n} \mathrm{E}(Y_r^2)\,\mathrm{E}(Y_s^2),$$

此即所求。（注意这部分用到了独立性和平均值为零，但没有用到各个 Y 同分布（identically distributed）。）

第 (iii) 部分

设每个 X_r 的平均值为 μ、标准差为 σ，并设 $S = X_1 + \cdots + X_n$。令 $Y_r = X_r - \mu$。各个 Y_r 互相独立且同分布，平均值为 0，而

$$\mathrm{E}(Y_r^2) = \sigma^2, \qquad \mathrm{E}(Y_r^4) = (\kappa + 3)\sigma^4,$$

第二式由移项 κ 的定义得出。于是 $T = \sum Y_r = S - n\mu$。由平移的结果，S 与 T 有相同的峰度，所以我们处理平均值为 0 的 T。

由独立性，$\mathrm{Var}(T) = \sum \mathrm{Var}(Y_r) = n\sigma^2$。第 (ii) 部分中的双重和共有 $\binom{n}{2} = \frac{1}{2}n(n-1)$ 项，每项都等于 σ^4，所以

$$\begin{aligned}\mathrm{E}(T^4) &= n(\kappa+3)\sigma^4 + 6\cdot\tfrac{1}{2}n(n-1)\sigma^4 \\ &= n\sigma^4(\kappa + 3 + 3(n-1)) = n\sigma^4(\kappa + 3n).\end{aligned}$$

因此 T 的峰度，亦即 S 的峰度，为

$$\frac{\mathrm{E}(T^4)}{(\mathrm{Var}\,T)^2} - 3 = \frac{n\sigma^4(\kappa+3n)}{n^2\sigma^4} - 3 = \frac{\kappa}{n} + 3 - 3 = \frac{\kappa}{n},$$

此即所求。

验算

- 取 $Y = \pm 1$，概率各为 $\frac{1}{2}$。则 $\mathrm{E}(Y^4) = \sigma^4 = 1$，所以 $\kappa = 1 - 3 = -2$。当 $n = 2$ 时，和的取值为 $-2, 0, 2$，概率分别为 $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$，所以 $\mathrm{E}(T^4) = 8$ 及 $\mathrm{Var}\,T = 2$。由此得峰度为 $8/4 - 3 = -1 = \kappa/2$，与预测相符。
- 当 $n = 1$ 时，第 (iii) 部分表示甚么都不变，理应如此。
- 由于互相独立的正态变量之和仍是正态变量，第 (iii) 部分取 $\kappa = 0$ 时必须再次得出 0，而事实确是如此。
- 我们用电脑以精确分数验证了：一个取三个值的变量，其最多六个独立副本之和的峰度恰好是 κ/n。我们亦以数值方法验证了，对多个 σ 值，$\int x^4 f(x)\,\mathrm{d}x = 3\sigma^4$。

常见错误

- 弄错 $Y_r^2 Y_s^2$ 的系数：每个无序对的系数是 $\binom{4}{2} = 6$（或每个有序对 3），不是 4 或 12。
- 在使用第 (ii) 部分之前忘记把 X_r 中心化，而第 (ii) 部分要求平均值为 0。
- 除以 σ^4 而不是 $(n\sigma^2)^2$：和的方差是 $n\sigma^2$。

STEP 之外

这里定义的峰度，是四阶累积量（cumulant）除以 σ^4。累积量是矩母函数（moment generating function）的对数展开式中的系数，而对于独立变量之和，它们直接相加。这就是为甚么和的标准化四阶累积量会像 $1/n$ 那样缩小，而三阶累积量（偏度（skewness））则像 $1/\sqrt{n}$ 那样缩小。这让我们定量地窥见中心极限定理（central limit theorem）：和看起来愈来愈像正态分布，而正态分布二阶以上的累积量全部为零。Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率导论》）在这个程度上讲解生成函数和中心极限定理。

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。