詳細解答

STEP III 2016，第 13 題

概率與統計 • 和的峰度

> **STEP III 2016，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/16-s3-q13

關鍵思路

峯度（kurtosis）只由偏離平均值的差構成，所以任何平移變量的操作都不會改變它；這讓我們可以把每個變量都中心化至 0，然後直接處理普通的矩（moments）。對於正態分佈（Normal distribution），四階矩可用分部積分法（integration by parts）求得，方法是分出一個 x 因子與指數函數配對。對於若干個互相獨立（independent）且平均值為零的變量之和，展開 T^4 再取期望值，凡是含有某個變量的一次方的項都會變成零。這就是第 (ii) 部分的全部內容，而它會直接用於第 (iii) 部分。

解答

平移。 設 $W = X - a$。由期望值的線性性質（linearity of expectation），$\mathrm{E}(W) = \mu - a$，所以

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

W 偏離其平均值的差，正好等於 X 偏離其平均值的差。因此 W 有相同的方差 $\mathrm{E}((X-\mu)^2) = \sigma^2$、相同的四階中心矩（central moment）$\mathrm{E}((X-\mu)^4)$，從而有相同的峯度。特別地，取 $a = \mu$，則每個隨機變量都與其中心化版本 $X - \mu$ 有相同的峯度，而後者的平均值為 0。

第 (i) 部分

設 $X \sim \mathrm{N}(0, \sigma^2)$，其概率密度函數（probability density function）為

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

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

想一想 既然 $\mathrm{e}^{-x^2/(2\sigma^2)}$ 沒有初等的原函數，我們怎樣積分 $x^4\mathrm{e}^{-x^2/(2\sigma^2)}$？

函數 $xf(x)$ 確實有原函數（antiderivative），就是 $-\sigma^2 f(x)$。所以寫成 $x^4 f(x) = x^3 \cdot xf(x)$，並取 $u = x^3$ 及 $\mathrm{d}v = xf(x)\,\mathrm{d}x$、$v = -\sigma^2 f(x)$ 進行分部積分：

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

邊界項為零，因為當 $x \to \pm\infty$ 時 $x^3\mathrm{e}^{-x^2/(2\sigma^2)} \to 0$（指數函數衰減的速度比任何冪的增長都快）。以 $u = x$ 重複同一步驟，得

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

因為 f 是概率密度函數，其積分為 1。（這也證實了方差是 σ^2。）因此 $\mathrm{E}(X^4) = 3\sigma^2 \cdot \sigma^2 = 3\sigma^4$，而

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

由平移的結果，每個正態分佈，無論其平均值是多少，峯度都是 0。

第 (ii) 部分

用多項式定理（multinomial theorem）展開 $T^4 = (Y_1 + \cdots + Y_n)^4$。每一項都是四個因子 Y_r 的乘積，而按照這四個因子如何分配到不同的下標，各項可分為五類：

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

其中不同的字母代表不同的下標。係數計算的是排列方式的數目：例如，$Y_r^2 Y_s^2$ 來自從 4 個括號中選出哪 2 個提供 Y_r，共有 $\binom{4}{2} = 6$ 種方法。每個無序對 $\{r, s\}$ 在 $Y_r^2 Y_s^2$ 這一類中只出現一次。

現在取期望值。各個 Y 互相獨立，所以不同 Y 的冪的乘積的期望值，等於各自期望值的乘積。在 $Y_r^3 Y_s$、$Y_r^2 Y_s Y_t$ 和 $Y_r Y_s Y_t Y_u$ 這幾類中，總有某個變量只以一次方出現，而它的因子 $\mathrm{E}(Y_s) = 0$ 令整個期望值為零。例如，

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

只有第一類和第三類保留下來，而同樣由獨立性，當 $r \neq s$ 時 $\mathrm{E}(Y_r^2 Y_s^2) = \mathrm{E}(Y_r^2)\,\mathrm{E}(Y_s^2)$。把各對寫成 $r < s$，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。（注意這部分用到了獨立性和平均值為零，但沒有用到各個 Y 同分佈（identically distributed）。）

第 (iii) 部分

設每個 X_r 的平均值為 μ、標準差為 σ，並設 $S = X_1 + \cdots + X_n$。令 $Y_r = X_r - \mu$。各個 Y_r 互相獨立且同分佈，平均值為 0，而

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

第二式由移項 κ 的定義得出。於是 $T = \sum Y_r = S - n\mu$。由平移的結果，S 與 T 有相同的峯度，所以我們處理平均值為 0 的 T。

由獨立性，$\mathrm{Var}(T) = \sum \mathrm{Var}(Y_r) = n\sigma^2$。第 (ii) 部分中的雙重和共有 $\binom{n}{2} = \frac{1}{2}n(n-1)$ 項，每項都等於 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 的峯度，亦即 S 的峯度，為

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

此即所求。

驗算

- 取 $Y = \pm 1$，概率各為 $\frac{1}{2}$。則 $\mathrm{E}(Y^4) = \sigma^4 = 1$，所以 $\kappa = 1 - 3 = -2$。當 $n = 2$ 時，和的取值為 $-2, 0, 2$，概率分別為 $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$，所以 $\mathrm{E}(T^4) = 8$ 及 $\mathrm{Var}\,T = 2$。由此得峯度為 $8/4 - 3 = -1 = \kappa/2$，與預測相符。
- 當 $n = 1$ 時，第 (iii) 部分表示甚麼都不變，理應如此。
- 由於互相獨立的正態變量之和仍是正態變量，第 (iii) 部分取 $\kappa = 0$ 時必須再次得出 0，而事實確是如此。
- 我們用電腦以精確分數驗證了：一個取三個值的變量，其最多六個獨立副本之和的峯度恰好是 κ/n。我們亦以數值方法驗證了，對多個 σ 值，$\int x^4 f(x)\,\mathrm{d}x = 3\sigma^4$。

常見錯誤

- 弄錯 $Y_r^2 Y_s^2$ 的係數：每個無序對的係數是 $\binom{4}{2} = 6$（或每個有序對 3），不是 4 或 12。
- 在使用第 (ii) 部分之前忘記把 X_r 中心化，而第 (ii) 部分要求平均值為 0。
- 除以 σ^4 而不是 $(n\sigma^2)^2$：和的方差是 $n\sigma^2$。

STEP 之外

這裏定義的峯度，是四階累積量（cumulant）除以 σ^4。累積量是矩母函數（moment generating function）的對數展開式中的係數，而對於獨立變量之和，它們直接相加。這就是為甚麼和的標準化四階累積量會像 $1/n$ 那樣縮小，而三階累積量（偏度（skewness））則像 $1/\sqrt{n}$ 那樣縮小。這讓我們定量地窺見中心極限定理（central limit theorem）：和看起來愈來愈像正態分佈，而正態分佈二階以上的累積量全部為零。Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率導論》）在這個程度上講解生成函數和中心極限定理。

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。