詳細解答

STEP II 1995，第 14 題

概率與統計 • 假設檢驗

STEP II 1995，第 14 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/95-s2-q14

關鍵思路

把密度寫成 $x^2\phi(x)$，其中 ϕ 是標準正態密度（standard normal density）。這樣一來，我們需要的每一個積分都變成標準正態變量 Z 的矩（moment），題目中關於 $\mathrm{E}(Z^4)$ 的提示正是用在這裏。要判斷數據是否「動搖我們的信心」，我們要問：假如我們的假設（hypothesis）（$\mu = 0$）成立，出現至少這樣極端的結果的概率有多大？如果這個概率極小，而把 μ 移動一下就能令數據變得平平無奇，那麼數據就對我們的假設不利。

解答

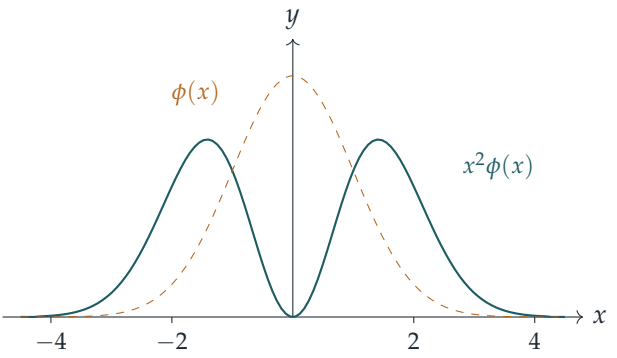
求 A

用分部積分法（integration by parts），取 $u = x$ 及 $\mathrm{d}v = x\mathrm{e}^{-x^2/2}\,\mathrm{d}x$，於是 $v = -\mathrm{e}^{-x^2/2}$：

$$\int_{-\infty}^{\infty} x^2\mathrm{e}^{-x^2/2}\,\mathrm{d}x = \Big[-x\mathrm{e}^{-x^2/2}\Big]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \mathrm{e}^{-x^2/2}\,\mathrm{d}x = 0 + \sqrt{2\pi},$$

這裏用了標準正態積分。總概率必須等於 1，所以

$$A = \frac{1}{\sqrt{2\pi}}, \qquad \text{即} \qquad \mathrm{f}(x) = x^2\phi(x), \quad \phi(x) = \tfrac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}.$$



這個密度是對稱的，在 $x = \pm\sqrt{2}$ 處有兩個峯，而在 $x = 0$ 處為零。由對稱性可知 $\mathrm{E}(X) = 0$，並且

$$\mathrm{E}(X^2) = \int_{-\infty}^{\infty} x^4\phi(x)\,\mathrm{d}x = \mathrm{E}(Z^4) = 3, \qquad \text{所以} \qquad \mathrm{Var}(X) = 3.$$

第 (i) 部分

在我們的假設之下，出現像 $a = 87.3$ 這樣偏遠的結果，可能性有多大？像之前一樣用分部積分法（這次 $v = -\phi(x)$，因為 $\phi'(x) = -x\phi(x)$），

$$\mathrm{P}(X \geq a) = \int_a^{\infty} x \cdot x\phi(x)\,\mathrm{d}x = a\phi(a) + \mathrm{P}(Z \geq a).$$

當 $a > 0$ 時，我們還有 $\mathrm{P}(Z \geq a) = \int_a^{\infty} \phi(x)\,\mathrm{d}x \leq \int_a^{\infty} \frac{x}{a}\phi(x)\,\mathrm{d}x = \frac{\phi(a)}{a}$，因為在積分範圍內 $x/a \geq 1$。因此，由對稱性，

$$\mathrm{P}(|X| \geq 87.3) \leq 2\left(a + \frac{1}{a}\right)\phi(a) \approx 70\,\mathrm{e}^{-3810.6} \approx 10^{-1653}.$$

這個數值小得荒謬：假如我們的假設成立，我們基本上永遠不會見到這樣的數值。另一方面，如果 μ 大約是 85 或 89（使得 87.3 正好落在 $\mathrm{f}(x - \mu)$ 的其中一個峯上），那麼 87.3 這個結果就完全平常。所以，單單一個 87.3 的結果，就應該徹底動搖我們的信心。唯一的出路是懷疑結果本身（例如記錄出錯），而明智的科學家在放棄假設之前，會先查證這一點。

? **想一想** 每一個個別數值的概率都是零，只憑一次觀測就下判斷，豈不是不公平？

重點不在於恰好取得 87.3 這個數值的概率，而在於取得一個離我們的假設所集中的位置至少這麼遠的數值的概率。同樣地，在我們的假設之下，87.3 處的密度大約是某個合適的備擇假設之下該處密度的 10^{-1650} 倍。無論從哪個角度看，觀測結果極端到這個程度，一次就綽綽有餘。

第 (ii) 部分

設 $X_1, \ldots, X_{1000}$ 為各個獨立的結果，$\bar{X}$ 為它們的平均值。在我們的假設之下，每個 X_i 的平均值為 0，方差（variance）為 3，所以

$$\mathrm{E}(\bar{X}) = 0, \qquad \mathrm{Var}(\bar{X}) = \frac{3}{1000}, \qquad \mathrm{sd}(\bar{X}) = \sqrt{0.003} \approx 0.0548.$$

方差是有限的，而 1000 又夠大，所以根據中心極限定理（central limit theorem），$\bar{X}$ 近似服從 $\mathrm{N}(0, 0.003)$，儘管 f 本身看起來一點也不像正態密度。觀測所得的平均值與 0 相距

$$\frac{0.23}{0.0548} \approx 4.2$$

個標準差（standard deviation），而

$$\mathrm{P}(|\bar{X}| \geq 0.23) \approx \mathrm{P}(|Z| \geq 4.2) \approx 3 \times 10^{-5}.$$

所以，假如我們的假設成立，把整個實驗重複十萬次，樣本平均值離 0 這麼遠的情況大約只會出現三次。而在 $\mu \approx 0.23$ 的備擇假設之下，同樣的數據就完全典型。這同樣應該嚴重動搖我們的信心，只是不及 (i) 那樣具壓倒性。在任何常用的顯著性水平（significance level）下，例如 1% 甚至 0.1%，我們都會否定 $\mu = 0$。

> **想一想** 我們可否改用切比雪夫不等式（Chebyshev's inequality），從而避開中心極限定理？

切比雪夫不等式給出 $\mathrm{P}(|\bar{X}| \geq 0.23) \leq 0.003/0.23^2 \approx 0.057$，當中不涉及任何近似。這個結果有參考價值，但不足以下定論，因為這樣只能在大約 6% 的水平上排除如此極端的結果。中心極限定理給出的數字要精確得多；而且有 1000 個觀測值，分佈的尾部又像 $\mathrm{e}^{-x^2/2}$ 那樣衰減，這個近似是可以信賴的。

驗算

- 辛普森法則（Simpson's rule）給出 $\int x^2 \mathrm{e}^{-x^2/2}\,\mathrm{d}x = 2.506628\ldots = \sqrt{2\pi}$，以及 $\int x^2 \mathrm{f}(x)\,\mathrm{d}x = 3.000$。
- 尾部公式 $\mathrm{P}(|X| \geq a) = 2(a\phi(a) + \mathrm{P}(Z \geq a))$ 在 $a = 3$ 時與數值積分的結果吻合（兩者皆為 0.029291）。
- $|X|$ 的密度，與由三個獨立 $\mathrm{N}(0,1)$ 變量組成的向量的長度的密度相同。用這個方法模擬 X，每次取 1000 個數值求平均，重複 4000 次，結果約有 4.9% 的平均值落在 2 sd 之外，接近正態分佈所預測的 4.6%，而且沒有一個超出 0.23。

常見錯誤

- 因為指數部分看似標準正態分佈，就取 $\mathrm{Var}(X) = 1$。額外的 x^2 把概率質量向外推，令方差變成 3。
- 在 (ii) 中，對單一個 X 而不是對 $\bar{X}$ 作正態檢驗，或者忘記把方差除以 1000。
- 在 (i) 中，辯稱「一次觀測證明不了甚麼」。結果極端到這個程度，一次觀測就足以說明問題。

STEP 之外

這是從基本原理出發看假設檢驗（hypothesis testing）。大學一年級的統計學課程會用 p 值（p-value）、似然比（likelihood ratio）（奈曼–皮爾遜引理，Neyman–Pearson lemma）和中心極限定理把它形式化。順帶一提，$|X|$ 服從自由度（degrees of freedom）為 3 的卡分佈（chi distribution），亦即氣體中分子速率的麥克斯韋–玻爾茲曼分佈（Maxwell–Boltzmann distribution）。Grimmett 與 Welsh 合著的 *Probability: An Introduction*（概率論入門教材）涵蓋了這裏用到的極限定理。

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。