详细解答

STEP II 1995，第 14 题

概率与统计 · 假设检验

STEP II 1995，第 14 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/95-s2-q14

关键思路

把密度写成 $x^2\phi(x)$，其中 ϕ 是标准正态密度（standard normal density）。这样一来，我们需要的每一个积分都变成标准正态变量 Z 的矩（moment），题目中关于 $\mathrm{E}(Z^4)$ 的提示正是用在这里。要判断数据是否“动摇我们的信心”，我们要问：假如我们的假设（hypothesis）（$\mu = 0$）成立，出现至少这样极端的结果的概率有多大？如果这个概率极小，而把 μ 移动一下就能令数据变得平平无奇，那么数据就对我们的假设不利。

解答

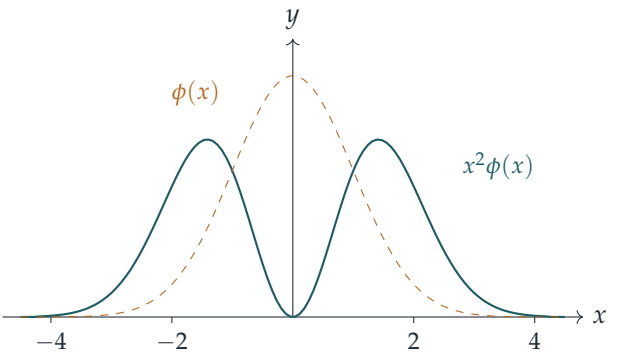
求 A

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

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

这里用了标准正态积分。总概率必须等于 1，所以

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



这个密度是对称的，在 $x = \pm\sqrt{2}$ 处有两个峰，而在 $x = 0$ 处为零。由对称性可知 $\mathrm{E}(X) = 0$，并且

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

第 (i) 部分

在我们的假设之下，出现像 $a = 87.3$ 这样偏远的结果，可能性有多大？像之前一样用分部积分法（这次 $v = -\phi(x)$，因为 $\phi'(x) = -x\phi(x)$），

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

当 $a > 0$ 时，我们还有 $\mathrm{P}(Z \geq a) = \int_a^{\infty} \phi(x)\,\mathrm{d}x \leq \int_a^{\infty} \frac{x}{a}\phi(x)\,\mathrm{d}x = \frac{\phi(a)}{a}$，因为在积分范围内 $x/a \geq 1$。因此，由对称性，

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

这个数值小得荒谬：假如我们的假设成立，我们基本上永远不会见到这样的数值。另一方面，如果 μ 大约是 85 或 89（使得 87.3 正好落在 $\mathrm{f}(x - \mu)$ 的其中一个峰上），那么 87.3 这个结果就完全平常。所以，单单一个 87.3 的结果，就应该彻底动摇我们的信心。唯一的出路是怀疑结果本身（例如记录出错），而明智的科学家在放弃假设之前，会先查证这一点。

想一想 每一个个别数值的概率都是零，只凭一次观测就下判断，岂不是不公平？

重点不在于恰好取得 87.3 这个数值的概率，而在于取得一个离我们的假设所集中的位置至少这么远的数值的概率。同样地，在我们的假设之下，87.3 处的密度大约是某个合适的备择假设之下该处密度的 10^{-1650} 倍。无论从哪个角度看，观测结果极端到这个程度，一次就绰绰有余。

第 (ii) 部分

设 $X_1, \ldots, X_{1000}$ 为各个独立的结果，$\bar{X}$ 为它们的平均值。在我们的假设之下，每个 X_i 的平均值为 0，方差（variance）为 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 又够大，所以根据中心极限定理（central limit theorem），$\bar{X}$ 近似服从 $\mathrm{N}(0, 0.003)$，尽管 f 本身看起来一点也不像正态密度。观测所得的平均值与 0 相距

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

个标准差（standard deviation），而

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

所以，假如我们的假设成立，把整个实验重复十万次，样本平均值离 0 这么远的情况大约只会出现三次。而在 $\mu \approx 0.23$ 的备择假设之下，同样的数据就完全典型。这同样应该严重动摇我们的信心，只是不及 (i) 那样具压倒性。在任何常用的显著性水平（significance level）下，例如 1% 甚至 0.1%，我们都会否定 $\mu = 0$。

> **想一想** 我们可否改用切比雪夫不等式（Chebyshev's inequality），从而避开中心极限定理?

切比雪夫不等式给出 $\mathrm{P}(|\bar{X}| \geq 0.23) \leq 0.003/0.23^2 \approx 0.057$，当中不涉及任何近似。这个结果有参考价值，但不足以下定论，因为这样只能在大约 6% 的水平上排除如此极端的结果。中心极限定理给出的数字要精确得多；而且有 1000 个观测值，分布的尾部又像 $\mathrm{e}^{-x^2/2}$ 那样衰减，这个近似是可以信赖的。

验算

- 辛普森法则（Simpson's rule）给出 $\int x^2 \mathrm{e}^{-x^2/2}\,\mathrm{d}x = 2.506628\ldots = \sqrt{2\pi}$，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 $a = 3$ 时与数值积分的结果吻合（两者皆为 0.029291）。
- $|X|$ 的密度，与由三个独立 $\mathrm{N}(0,1)$ 变量组成的向量的长度的密度相同。用这个方法模拟 X，每次取 1000 个数值求平均，重复 4000 次，结果约有 4.9% 的平均值落在 2 sd 之外，接近正态分布所预测的 4.6%，而且没有一个超出 0.23。

常见错误

- 因为指数部分看似标准正态分布，就取 $\mathrm{Var}(X) = 1$。额外的 x^2 把概率质量向外推，令方差变成 3。
- 在 (ii) 中，对单一个 X 而不是对 $\bar{X}$ 作正态检验，或者忘记把方差除以 1000。
- 在 (i) 中，辩称“一次观测证明不了甚么”。结果极端到这个程度，一次观测就足以说明问题。

STEP 之外

这是从基本原理出发看假设检验（hypothesis testing）。大学一年级的统计学课程会用 p 值（p-value）、似然比（likelihood ratio）（奈曼–皮尔逊引理，Neyman–Pearson lemma）和中心极限定理把它形式化。顺带一提，$|X|$ 服从自由度（degrees of freedom）为 3 的卡分布（chi distribution），亦即气体中分子速率的麦克斯韦–玻尔兹曼分布（Maxwell–Boltzmann distribution）。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门教材）涵盖了这里用到的极限定理。

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。