详细解答

# STEP II 2009，第 12 题

概率与统计 • 半正态分布

**STEP II 2009，第 12 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/09-s2-q12

## 关键思路

指数 $-2x^2$ 其实是一个乔装了的正态分布密度函数（normal density）。把它写成 $-2x^2 = -\dfrac{x^2}{2\sigma^2}$，便得 $\sigma = \frac{1}{2}$，所以 $f$ 恰好是 $Y \sim \mathrm{N}(0, \frac{1}{4})$ 的密度函数的两倍，并在 0 处截断。换句话说，$X$ 与 $|Y|$ 有相同的分布。一旦看出这一点，归一化常数（normalising constant）、$\mathrm{E}(X^2)$ 和中位数（median）全都可以直接由正态分布的性质得出。只有 $\mathrm{E}(X)$ 需要计算一个（非常容易的）积分。

## 解答

以下一律设 $Y \sim \mathrm{N}(0, \frac{1}{4})$ 及 $Z \sim \mathrm{N}(0,1)$，于是 $Y$ 与 $\frac{1}{2}Z$ 有相同的分布，而它的密度函数为

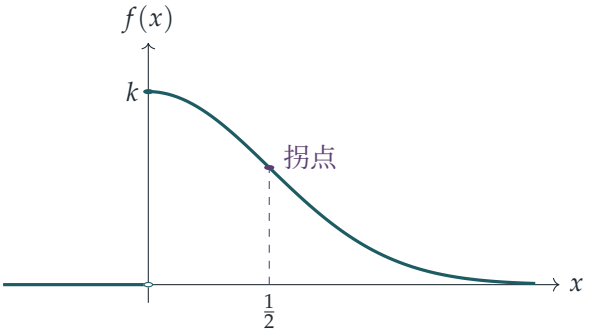
$$g(y) = \frac{1}{\frac{1}{2}\sqrt{2\pi}}\,\mathrm{e}^{-y^2/(2\cdot\frac{1}{4})} = \frac{2}{\sqrt{2\pi}}\,\mathrm{e}^{-2y^2}.$$

### 第 (i) 部分

当 $x \geq 0$ 时，$f'(x) = -4kx\,\mathrm{e}^{-2x^2} \leq 0$，所以 $f$ 由 $x = 0$ 处的最大值 $k$ 开始递减，而该处的斜率为零。另外

$$f''(x) = k(16x^2 - 4)\,\mathrm{e}^{-2x^2},$$

它在 $x = \frac{1}{2}$ 处变号，所以在 $x = \frac{1}{2}$ 处有一个拐点（point of inflection）。当 $x \to \infty$ 时，$f(x) \to 0$，而且趋近得非常快；当 $x < 0$ 时，$f(x) = 0$。图像是钟形曲线的右半部分（由第 (ii) 部分可知 $k \approx 1.60$），并在 $x = 0$ 处由 0 跳升至 $k$。




**第 (ii) 部分**

总概率必须等于 1。由于 $g$ 关于 0 对称，而且在整条实数线上的积分为 1，所以它在 $[0,\infty)$ 上的积分为 $\frac{1}{2}$：

$$\int_0^\infty \frac{2}{\sqrt{2\pi}}\,\mathrm{e}^{-2x^2}\,\mathrm{d}x = \frac{1}{2} \quad\Longrightarrow\quad \int_0^\infty \mathrm{e}^{-2x^2}\,\mathrm{d}x = \frac{\sqrt{2\pi}}{4}.$$

因此 $k\cdot\dfrac{\sqrt{2\pi}}{4} = 1$，得

$$k = \frac{4}{\sqrt{2\pi}} = 2\sqrt{\frac{2}{\pi}} \approx 1.596.$$

留意当 $x \geq 0$ 时 $f(x) = 2g(x)$，这印证了 $X$ 的分布与 $|Y|$ 相同。

**第 (iii) 部分**

求平均值时，因子 $x$ 令积分变得十分简单，因为 $x\,\mathrm{e}^{-2x^2}$ 是 $\mathrm{e}^{-2x^2}$ 的导数的常数倍：

$$\mathrm{E}(X) = k\int_0^\infty x\,\mathrm{e}^{-2x^2}\,\mathrm{d}x = k\Big[-\tfrac{1}{4}\mathrm{e}^{-2x^2}\Big]_0^\infty = \frac{k}{4} = \frac{1}{\sqrt{2\pi}} \approx 0.399.$$

? **想一想** 怎样才能不用冗长的分部积分法（integration by parts）而求出 $\displaystyle\int_0^\infty x^2\,\mathrm{e}^{-2x^2}\,\mathrm{d}x$?

再次利用正态分布。函数 $x^2g(x)$ 是偶函数（even function），所以

$$\mathrm{E}(X^2) = \int_0^\infty x^2\cdot 2g(x)\,\mathrm{d}x = \int_{-\infty}^\infty x^2 g(x)\,\mathrm{d}x = \mathrm{E}(Y^2) = \mathrm{Var}(Y) = \frac{1}{4},$$

因为 $\mathrm{E}(Y) = 0$。（取 $u = x$、$\mathrm{d}v = x\mathrm{e}^{-2x^2}\,\mathrm{d}x$ 作分部积分，亦得相同结果：$\int_0^\infty x^2\mathrm{e}^{-2x^2}\,\mathrm{d}x = \frac{1}{4}\int_0^\infty \mathrm{e}^{-2x^2}\,\mathrm{d}x = \frac{1}{4k}$。）因此

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = \frac{1}{4} - \frac{1}{2\pi} = \frac{\pi-2}{4\pi} \approx 0.0908.$$

**第 (iv) 部分**

中位数 $m$ 满足 $\mathrm{P}(X \le m) = \frac{1}{2}$。由于 $X$ 的分布与 $|Y|$ 相同，而 $Y$ 的分布与 $\frac{1}{2}Z$ 相同，

$$\mathrm{P}(X \le m) = \mathrm{P}(|Z| \le 2m) = 2\Phi(2m) - 1,$$

这里用到了标准正态分布（standard normal distribution）的对称性。（直接计算：$\mathrm{P}(X \le m) = \int_0^m 2g(x)\,\mathrm{d}x = 2(\mathrm{P}(Y \le m) - \frac{1}{2})$。）令上式等于 $\frac{1}{2}$，得

$$\Phi(2m) = \frac{3}{4}.$$

查表得 $\Phi^{-1}(0.75) = 0.6745$，所以 $2m = 0.6745$，而

$$m \approx 0.337 \text{ （准确至三位有效数字）}.$$

## 验算

取 $k = 2\sqrt{2/\pi}$，以数值方法（辛普森法则，Simpson's rule）在 $[0, 10]$ 上对 $f$ 积分，得总概率为 1.0000、$\mathrm{E}(X) = 0.39894$ 及 $\mathrm{Var}(X) = 0.090845$，与 $1/\sqrt{2\pi}$ 和 $(\pi - 2)/(4\pi)$ 吻合。以数值方法解 $\int_0^m f(x)\,\mathrm{d}x = \frac{1}{2}$，得 $m = 0.33724$。以 200 000 个标准差（standard deviation）为 $\frac{1}{2}$ 的正态随机数模拟 $|Y|$，得平均值 0.399 及方差（variance）0.091，并有 50.0% 的数值小于 0.337。合理性检查：中位数 0.337 小于平均值 0.399，对于一个右尾较长的分布来说，理应如此。

## 常见错误

- 从指数误取 $\sigma^2 = 2$ 或 $\sigma = \frac{1}{4}$：应把 $2x^2$ 与 $x^2/(2\sigma^2)$ 对照，得 $\sigma^2 = \frac{1}{4}$。
- 忘记因子 2：在半条实数线上，正态密度函数的积分只有 $\frac{1}{2}$。
- 解 $\Phi(2m) = \frac{1}{2}$，结果得出 $m = 0$。$X$ 在 $m$ 以下的那一半概率，对应的是 $Y$ 中间的那一半概率，所以条件应为 $\Phi(2m) = \frac{3}{4}$。

## STEP 之外

$X$ 是一个半正态（half-normal）随机变量（random variable），而当 $Y$ 服从正态分布时，$|Y|$ 就是随机变量的变换（transformation of a random variable）最简单的例子。一般而言，若 $X = h(Y)$，则 $X$ 的密度函数可由 $Y$ 的密度函数对所有原像（preimage）求和而得，这正是此处因子 2 的由来。同样的想法可以证明 $Z^2$ 服从自由度（degree of freedom）为一的卡方分布（chi-squared distribution），而这正是许多统计学内容的起点。Grimmett

and Welsh, *Probability: An Introduction*（概率入门教材）对随机变量的函数有仔细的论述。

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。