详细解答

STEP I 2011，第 13 题

概率与统计 · 连续随机变量

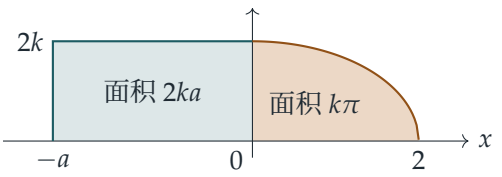
STEP I 2011，第 13 题

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/11-s1-q13

关键思路

先画出密度函数（density function）的图像。它由两部分拼成：在 $[-a, 0)$ 上是一个高为 $2k$ 的长方形，在 $[0, 2]$ 上则是一个半径为 2 的四分之一圆（垂直方向按比例 k 缩放）。图像下方的面积就是概率，因此单靠长方形和四分之一圆，大部分工作已经完成，毋须做任何积分。总面积必须等于 1，由此便可定出 k。之后所有关于 d 的问题，都归结为同一件事：某一点右方有多少面积?



解答

求 k

在 $[0, 2]$ 上，$\sqrt{4 - x^2}$ 的图像是圆 $x^2 + y^2 = 4$ 的四分之一，所以

$$\int_0^2 \sqrt{4 - x^2}\,\mathrm{d}x = \tfrac{1}{4}\pi \cdot 2^2 = \pi.$$

（用题目给出的原函数（antiderivative）计算，结果一致：$2\arcsin 1 + 0 - 0 = \pi$。）因此总概率为 $2ka + k\pi = 1$，即

$$k = \frac{1}{2a + \pi}.$$

第 (i) 部分

把 $\mathrm{E}(X)$ 的积分在 0 处分拆：

$$\int_{-a}^{0} 2kx\,\mathrm{d}x = k[x^2]_{-a}^{0} = -ka^2.$$

至于第二部分，$(4-x^2)^{3/2}$ 的导数是 $-3x(4-x^2)^{1/2}$，所以

$$\int_{0}^{2} kx\sqrt{4-x^2}\,\mathrm{d}x = k\Big[-\tfrac{1}{3}(4-x^2)^{3/2}\Big]_{0}^{2} = k\cdot\tfrac{1}{3}\cdot 8 = \tfrac{8}{3}k.$$

两部分相加，得

$$\mathrm{E}(X) = k\Big(\frac{8}{3} - a^2\Big) = \frac{8-3a^2}{3(2a+\pi)}.$$

第 (ii) 部分

0 右方的概率就是四分之一圆的面积：

$$\mathrm{P}(X>0) = k\pi = \frac{\pi}{2a+\pi}.$$

> ? **想一想** 知道了 $\mathrm{P}(X>0)$，为何就能判断 d 的正负？

由于密度函数在 $(-a,2)$ 上为正，函数 $t \mapsto \mathrm{P}(X>t)$ 在 $-a \le t \le 2$ 上连续（continuous）且严格递减（strictly decreasing）。它由 1 降至 0，所以恰好有一个 d 满足 $\mathrm{P}(X>d) = \frac{1}{10}$。由于函数递减，$d<0$ 当且仅当 $\mathrm{P}(X>0) < \frac{1}{10}$。换句话说，当 0 右方的概率不足十分之一时，$d$ 便是负数。

现在假设 $2a > 9\pi$，则 $2a+\pi > 10\pi$，所以

$$\mathrm{P}(X>0) = \frac{\pi}{2a+\pi} < \frac{1}{10},$$

因此 $d<0$，得证。（这个论证反过来亦成立：只有当 $2a>9\pi$ 时才有 $d<0$。）

在这情况下 $-a \le d < 0$，所以 d 右方的面积，就是长方形由 d 到 0、宽度为 $-d$ 的那一部分，再加上整个四分之一圆：

$$\mathrm{P}(X>d) = 2k(-d) + k\pi = \frac{\pi-2d}{2a+\pi}.$$

令它等于 $\frac{1}{10}$，得 $10\pi - 20d = 2a+\pi$，所以

$$d = \frac{9\pi-2a}{20}.$$

这个值确实位于 $[-a,0)$ 之内：因为 $2a>9\pi$，所以它是负数；而 $d \ge -a$ 等价于 $9\pi+18a \ge 0$，这显然成立。

第 (iii) 部分

现在 $d=\sqrt{2}>0$，所以 d 落在圆弧部分之下，而

$$\mathrm{P}(X>\sqrt{2})=k\int_{\sqrt{2}}^{2}\sqrt{4-x^2}\,\mathrm{d}x.$$

利用题目给出的结果，

$$\begin{aligned}\int_{\sqrt{2}}^{2}\sqrt{4-x^2}\,\mathrm{d}x&=\left[2\arcsin(\tfrac{1}{2}x)+\tfrac{1}{2}x\sqrt{4-x^2}\right]_{\sqrt{2}}^{2}\\&=(2\cdot\tfrac{\pi}{2}+0)-(2\cdot\tfrac{\pi}{4}+\tfrac{\sqrt{2}}{2}\cdot\sqrt{2})\\&=\frac{\pi}{2}-1.\end{aligned}$$

（从几何角度看，这是被 $x=\sqrt{2}$ 处的弦（chord）割出的弓形（circular segment），它等于圆面积的八分之一，即 $\frac{\pi}{2}$，减去一个面积为 1 的直角三角形。）所以

$$\frac{\frac{\pi}{2}-1}{2a+\pi}=\frac{1}{10}\quad\Longrightarrow\quad 5\pi-10=2a+\pi\quad\Longrightarrow\quad a=2\pi-5.$$

这是正数（$a\approx 1.283$），并且满足 $2a\le 9\pi$；由于 $d>0$，这与第 (ii) 部分的结论吻合。

验算

- 在第 (i) 部分，取 $a=1$，均值为 $\frac{5}{3(2+\pi)}\approx 0.3242$。以 2×10^5 个样本进行蒙地卡罗模拟（Monte Carlo simulation），得出 0.3236。
- 在临界情况 $2a=9\pi$ 下，第 (ii) 部分的公式给出 $d=0$，与条件一致。取 $a=20$，公式给出 $d=\frac{9\pi-40}{20}\approx -0.586$，而以数值方法解 $\mathrm{P}(X>d)=\frac{1}{10}$ 亦得出相同的值。
- 在第 (iii) 部分，取 $a=2\pi-5$，以辛普森法则（Simpson's rule）计算得 $\mathrm{P}(X>\sqrt{2})=0.1000$。

常见错误

- 没有求出 k。归一化（normalisation）条件 $2ka+k\pi=1$ 是隐藏的第一步，而所有答案都只应以 a 表示。
- 在第 (ii) 部分漏掉四分之一圆，只写 $\mathrm{P}(X>d)=-2kd$。其实 $[0,2]$ 上的全部概率，同样位于负数 d 的右方。
- 把 $\int_{-a}^{0}2kx\,\mathrm{d}x$ 的正负号弄错。由于整个区间上 $x<0$，这个积分是负数。

STEP 之外

d 这个值就是 X 的上 10% 点，即 0.9 分位数（quantile）。要求出它，就要求累积分布函数（cumulative distribution function）F 的反函数；而这个反函数，正正是用均匀分布的随机变量去模拟另一个随机变量时所用的工具（逆变换抽样（inverse transform sampling））。这个密度函数形状混合，一边平坦、一边弯曲，说明了为何分位数往往要分段求出。Grimmett 和 Welsh 所著的 *Probability: An Introduction*（《概率导论》）对分布函数和分位数有仔细的论述。

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。