详细解答

STEP II 2007，第 14 题

概率与统计 · 分段概率密度函数

> **STEP II 2007，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/07-s2-q14

关键思路

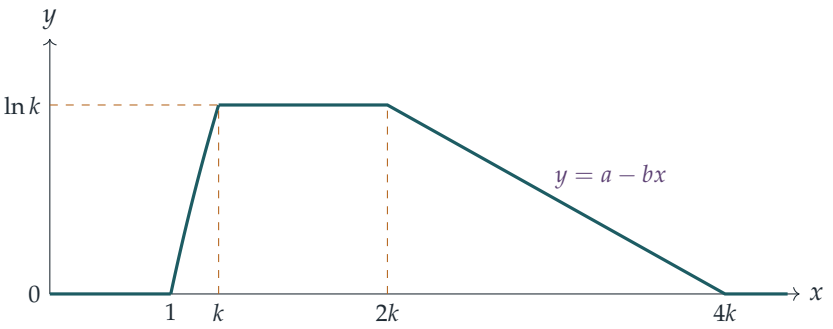
“连续”这两个字举足轻重。分段定义的密度函数（density function）在每个分界点都必须衔接起来，由此我们马上得到关于 a 和 b 的两条方程。余下的一个条件，即图像下的总面积为 1，便可确定 k。至于中位数（median），我们逐段把面积累加，直至超过二分之一为止。

解答

全题之中 $k > 1$：否则区间 $1 \le x \le k$ 便是空的，而且无论如何我们都需要 $\ln k > 0$，密度函数才会是非负，并且在 $x = 1$ 之后不恒等于零。

第 (i) 部分

在 $[1, k]$ 上，图像是常见的对数曲线，在 $x = 1$ 时由 0 开始（因此与取值为零的一段连续地衔接），递增而且是凹的（concave）。它在 $x = k$ 时到达高度 $\ln k$，保持在 $\ln k$ 的水平直至 $x = 2k$，然后沿直线下降，在 $x = 4k$ 时降至 0（下文会证明连续性恰好迫使如此），其后一直为零。草图采用了我们在第 (ii) 部分求得的数值：$k = \mathrm{e}^{1/3} \approx 1.40$，因此平台的高度为 $\frac{1}{3}$。



第 (ii) 部分

连续性（*continuity*）。在 $x = k$ 处，两条公式都给出 $\ln k$，所以这里没有新的资料。在 $x = 2k$ 和 $x = 4k$ 处，连续性要求

$$a - 2bk = \ln k, \qquad a - 4bk = 0.$$

两式相减，得 $2bk = \ln k$，所以

$$b = \frac{\ln k}{2k}, \qquad a = 4bk = 2\ln k.$$

总概率。我们把面积分成三部分。用分部积分法（integration by parts）（取 $u = \ln x$, $\mathrm{d}v = \mathrm{d}x$），

$$\int_1^k \ln x \,\mathrm{d}x = \Big[x\ln x - x\Big]_1^k = k\ln k - k + 1.$$

平台是一个阔 k、高 $\ln k$ 的长方形，面积为 $k\ln k$。最后一部分是一个三角形，底为 $4k - 2k = 2k$，高为 $\ln k$，面积为 $\frac{1}{2}\cdot 2k\cdot \ln k = k\ln k$。所以

$$1 = (k\ln k - k + 1) + k\ln k + k\ln k = 3k\ln k - k + 1,$$

化简后得

$$k(3\ln k - 1) = 0.$$

由于 $k > 1$，不可能有 $k = 0$，所以 $\ln k = \frac{1}{3}$，而

$$k = \mathrm{e}^{1/3} \approx 1.396, \qquad a = 2\ln k = \tfrac{2}{3}, \qquad b = \frac{\ln k}{2k} = \frac{1}{6\,\mathrm{e}^{1/3}} \approx 0.119.$$

留意 $b > 0$，所以最后一段确实如图所示向下倾斜。

第 (iii) 部分

我们把概率逐段累加，找出中位数 m 落在哪一段。利用 $\ln k = \frac{1}{3}$，

$$\mathrm{P}(X \le k) = k\ln k - k + 1 = 1 - \tfrac{2k}{3} \approx 0.070,$$

而平台再加上 $k\ln k = \frac{k}{3}$，所以

$$\mathrm{P}(X \le 2k) = 1 - \tfrac{2k}{3} + \tfrac{k}{3} = 1 - \tfrac{k}{3} \approx 0.535.$$

想一想 中位数落在图像的哪一段上?

由于 $0.070 < \frac{1}{2} < 0.535$，中位数落在平台上，即 $k \le m \le 2k$。在这一段，密度函数恒等于 $\frac{1}{3}$，所以

$$\mathrm{P}(X \le m) = 1 - \tfrac{2k}{3} + \tfrac{1}{3}(m-k) = \tfrac{1}{2}.$$

乘以 3：$3 - 2k + m - k = \frac{3}{2}$，所以

$$m = 3k - \tfrac{3}{2} = 3\mathrm{e}^{1/3} - \tfrac{3}{2} \approx 2.69.$$

这个值确实介乎 $k \approx 1.40$ 与 $2k \approx 2.79$ 之间，正如所应有的那样。

验算

- 取 $k = \mathrm{e}^{1/3}$、$a = \frac{2}{3}$、$b = 1/(6\mathrm{e}^{1/3})$，对密度函数作数值积分（中点法则），得总面积为 1.0000000，而 $\int_1^{2.68684} f(x)\,\mathrm{d}x = 0.5000000$。
- 中位数的合理性检查：略多于一半的概率落在 $2k$ 的左边（约 0.535），所以中位数应该在 $2k \approx 2.79$ 稍左的位置；果然 $m \approx 2.69$。

常见错误

- 忘记了 $\int_1^k \ln x\,\mathrm{d}x$ 的下限：$-[x\ln x - x]$ 在 $x = 1$ 处给出的 $+1$，正是令关于 k 的方程可以如此俐落地因式分解的原因。
- 因为中位数在支集（support）$[1, 4k]$ 的“中间”，便假设它落在倾斜的一段上；务必先把面积累加。

STEP 之外

本题的密度函数由几段拼合而成，而各常数则由连续性和归一化（normalisation）确定。到了大学，你会遇到*累积分布函数*（cumulative distribution function）$F(x) = \mathrm{P}(X \le x)$，并以它作为最基本的对象：每一个随机变量（random variable），不论是离散的还是连续的，都有累积分布函数，而中位数这类量不过是 $F(m) = \frac{1}{2}$ 的解。Grimmett and Welsh, *Probability: An Introduction*（概率论入门教材）清楚地阐述了这个观点。

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。