詳細解答

STEP II 2007，第 14 題

概率與統計 • 分段概率密度函數

STEP II 2007，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/07-s2-q14

關鍵思路

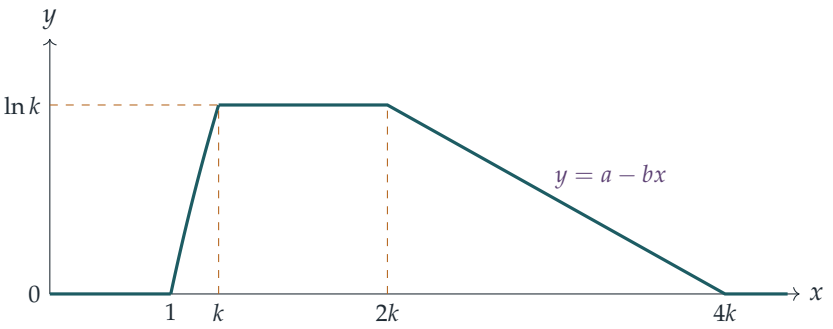
「連續」這兩個字舉足輕重。分段定義的密度函數（density function）在每個分界點都必須銜接起來，由此我們馬上得到關於 a 和 b 的兩條方程。餘下的一個條件，即圖像下的總面積為 1，便可確定 k。至於中位數（median），我們逐段把面積累加，直至超過二分之一為止。

解答

全題之中 $k > 1$：否則區間 $1 \le x \le k$ 便是空的，而且無論如何我們都需要 $\ln k > 0$，密度函數才會是非負，並且在 $x = 1$ 之後不恆等於零。

第 (i) 部分

在 $[1, k]$ 上，圖像是常見的對數曲線，在 $x = 1$ 時由 0 開始（因此與取值為零的一段連續地銜接），遞增而且是凹的（concave）。它在 $x = k$ 時到達高度 $\ln k$，保持在 $\ln k$ 的水平直至 $x = 2k$，然後沿直線下降，在 $x = 4k$ 時降至 0（下文會證明連續性恰好迫使如此），其後一直為零。草圖採用了我們在第 (ii) 部分求得的數值：$k = \mathrm{e}^{1/3} \approx 1.40$，因此平台的高度為 $\frac{1}{3}$。



第 (ii) 部分

連續性（*continuity*）。在 $x = k$ 處，兩條公式都給出 $\ln k$，所以這裏沒有新的資料。在 $x = 2k$ 和 $x = 4k$ 處，連續性要求

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

兩式相減，得 $2bk = \ln k$，所以

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

總概率。我們把面積分成三部分。用分部積分法（integration by parts）（取 $u = \ln x$，$\mathrm{d}v = \mathrm{d}x$），

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

平台是一個闊 k、高 $\ln k$ 的長方形，面積為 $k\ln k$。最後一部分是一個三角形，底為 $4k - 2k = 2k$，高為 $\ln k$，面積為 $\frac{1}{2}\cdot 2k\cdot \ln k = k\ln k$。所以

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

化簡後得

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

由於 $k > 1$，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 $b > 0$，所以最後一段確實如圖所示向下傾斜。

第 (iii) 部分

我們把概率逐段累加，找出中位數 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 中位數落在圖像的哪一段上？

由於 $0.070 < \frac{1}{2} < 0.535$，中位數落在平台上，即 $k \le m \le 2k$。在這一段，密度函數恆等於 $\frac{1}{3}$，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

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

這個值確實介乎 $k \approx 1.40$ 與 $2k \approx 2.79$ 之間，正如所應有的那樣。

驗算

- 取 $k = \mathrm{e}^{1/3}$、$a = \frac{2}{3}$、$b = 1/(6\mathrm{e}^{1/3})$，對密度函數作數值積分（中點法則），得總面積為 1.0000000，而 $\int_1^{2.68684} f(x)\,\mathrm{d}x = 0.5000000$。
- 中位數的合理性檢查：略多於一半的概率落在 $2k$ 的左邊（約 0.535），所以中位數應該在 $2k \approx 2.79$ 稍左的位置；果然 $m \approx 2.69$。

常見錯誤

- 忘記了 $\int_1^k \ln x\,\mathrm{d}x$ 的下限：$-[x\ln x - x]$ 在 $x = 1$ 處給出的 $+1$，正是令關於 k 的方程可以如此俐落地因式分解的原因。
- 因為中位數在支集（support）$[1, 4k]$ 的「中間」，便假設它落在傾斜的一段上；務必先把面積累加。

STEP 之外

本題的密度函數由幾段拼合而成，而各常數則由連續性和歸一化（normalisation）確定。到了大學，你會遇到*累積分佈函數*（cumulative distribution function）$F(x) = \mathrm{P}(X \le x)$，並以它作為最基本的對象：每一個隨機變量（random variable），不論是離散的還是連續的，都有累積分佈函數，而中位數這類量不過是 $F(m) = \frac{1}{2}$ 的解。Grimmett and Welsh, *Probability: An Introduction*（概率論入門教材）清楚地闡述了這個觀點。

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。