詳細解答

STEP II 2010，第 12 題

概率與統計 • 平均值與中位數

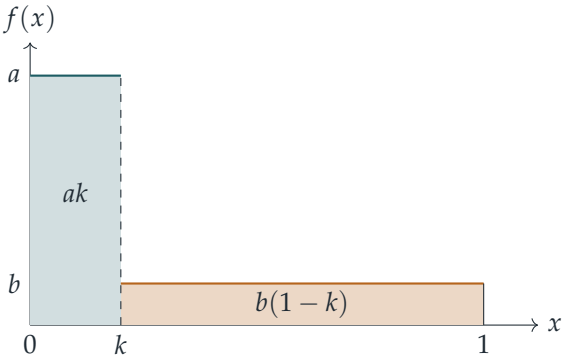
STEP II 2010，第 12 題

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/10-s2-q12

關鍵思路

這個密度函數（density function）是只有兩級的階梯函數（step function），所以每個概率都是一個長方形的面積，而唯一真正的條件就是總面積等於 1。這個條件把 k 用 a 和 b 表示出來；一旦消去 k，餘下的全是代數運算。至於中位數（median），問題只在於那個較高的第一個長方形是否已經佔了一半的概率。



解答

圖中兩個長方形的面積分別是 ak 和 $b(1-k)$，所以密度函數的積分等於 1，當且僅當

$$ak + b(1-k) = 1, \qquad \text{即} \qquad k = \frac{1-b}{a-b}.$$

（由於 $a > b$，我們可以除以 $a-b$。）以下一直採用這個 k 值。

為何 $a > 1$ 且 $b < 1$。 方程 $ak + b(1-k) = 1$ 表示 1 是 a 和 b 的加權平均（weighted average），權重為 k 和 $1-k$，兩者都嚴格介乎 0 與 1 之間。權重為正的加權平均必定嚴格介乎兩個數值之間，所以 $b < 1 < a$。不用文字來說，就是：$1 - b = k(a-b) > 0$ 及 $a - 1 = (1-k)(a-b) > 0$。

第 (i) 部分

在每一段上對 $xf(x)$ 積分，

$$\mathrm{E}(X) = \int_0^k ax\,\mathrm{d}x + \int_k^1 bx\,\mathrm{d}x = \tfrac{1}{2}ak^2 + \tfrac{1}{2}b(1-k^2) = \tfrac{1}{2}(b + (a-b)k^2).$$

現在 $(a-b)k^2 = \dfrac{(1-b)^2}{a-b}$，所以

$$\mathrm{E}(X) = \frac{b(a-b) + (1-b)^2}{2(a-b)} = \frac{ab - b^2 + 1 - 2b + b^2}{2(a-b)} = \frac{1 - 2b + ab}{2(a-b)},$$

正如所求。

第 (ii) 部分

累積分佈函數（cumulative distribution function）是

$$\mathrm{F}(x) = \begin{cases} ax & 0 \le x < k, \\ ak + b(x-k) & k \le x \le 1. \end{cases}$$

它在 $[0,1]$ 上連續而且嚴格遞增，所以中位數 M 是 $\mathrm{F}(M) = \frac{1}{2}$ 的唯一解。該用哪一條公式，取決於第一個長方形是否佔了至少一半的概率。

> **想一想** 怎樣把「第一個長方形的面積至少是 $\frac{1}{2}$」化為關於 a 和 b 的條件？

代入 k 的值，並消去正數分母 $a-b$：

$$ak \ge \tfrac{1}{2} \iff 2a(1-b) \ge a - b \iff 2a - 2ab \ge a - b \iff a + b \ge 2ab.$$

情況 $a + b \ge 2ab$。這時 $\mathrm{F}(k) = ak \ge \frac{1}{2}$，所以中位數位於 $[0,k]$ 之內，並滿足 $aM = \frac{1}{2}$：

$$M = \frac{1}{2a}.$$

情況 $a + b \le 2ab$。這時 $ak \le \frac{1}{2}$，所以中位數位於 $[k,1]$ 之內。從右端入手最為簡便：M 右方的概率是 $b(1-M)$，而它必須等於 $\frac{1}{2}$。因此

$$M = 1 - \frac{1}{2b} = \frac{2b-1}{2b}.$$

當 $a + b = 2ab$ 時，兩條公式都適用，而且結果一致，因為這時 $1 - \frac{1}{2b} = \frac{2ab-a}{2ab} = \frac{b}{2ab} = \frac{1}{2a}$。

第 (iii) 部分

我們在每個情況下計算 $\mathrm{E}(X) - M$，並證明它是正數。

情況 $a + b \geq 2ab$。通分，以 $2a(a-b)$ 為公分母，

$$\mathrm{E}(X) - \frac{1}{2a} = \frac{a(1-2b+ab) - (a-b)}{2a(a-b)} = \frac{b - 2ab + a^2 b}{2a(a-b)} = \frac{b(a-1)^2}{2a(a-b)}.$$

情況 $a + b \leq 2ab$。通分，以 $2b(a-b)$ 為公分母，

$$\begin{aligned}\mathrm{E}(X) - \frac{2b-1}{2b} &= \frac{b(1-2b+ab) - (2b-1)(a-b)}{2b(a-b)} \\ &= \frac{(b - 2b^2 + ab^2) - (2ab - 2b^2 - a + b)}{2b(a-b)} = \frac{a(1-b)^2}{2b(a-b)}.\end{aligned}$$

想一想 為甚麼這些數值是嚴格為正，而不僅僅是非負？

在兩個情況下，a、b 和 $a-b$ 都是正數，而平方項嚴格為正，因為我們一開始已證明 $a > 1$ 且 $b < 1$。所以在任何情況下都有 $\mathrm{E}(X) - M > 0$，即 $M < \mathrm{E}(X)$。

驗算

- 取 $a = 3$、$b = \frac{1}{2}$：這時 $k = \frac{1}{5}$，$a + b = \frac{7}{2} \geq 3 = 2ab$，而 $\mathrm{E}(X) = \frac{1-1+\frac{3}{2}}{5} = 0.3$，$M = \frac{1}{6}$。兩者之差 $0.1\overline{3}$ 與 $\frac{b(a-1)^2}{2a(a-b)} = \frac{2}{15}$ 吻合。
- 取 $a = \frac{5}{4}$、$b = \frac{9}{10}$：這次 $a + b = 2.15 \leq 2.25 = 2ab$，所以 $M = 1 - \frac{5}{9} = \frac{4}{9} \approx 0.444$，而 $\mathrm{E}(X) = \frac{13}{28} \approx 0.464$。
- 當 $a \to 1^+$（因而 $b \to 1^-$）時，密度函數趨向均勻分佈（uniform distribution）的密度函數，兩個差值都趨向 0：平均值（mean）與中位數同為 $\frac{1}{2}$，按對稱性理應如此。

我們以 300 對隨機選取的 (a, b) 作了數值檢驗：對 F 使用二分法（bisection）求得的中位數與分情況的公式吻合，以中點求積法（midpoint quadrature）求得的平均值與第 (i) 部分吻合，而 $\mathrm{E}(X) - M$ 亦等於上面已因式分解的表達式。

常見錯誤

- 忘記消去 k：答案須只用 a 和 b 表示。
- 把第二個中位數寫成 $b(M-k) = \frac{1}{2}$ 的解，忘記了在 $[0, k)$ 上已經累積的概率 ak。
- 只得出 $\mathrm{E}(X) \geq M$ 的結論：嚴格不等式需要 $a \neq 1$ 和 $b \neq 1$，而這正是題目開首部分所提供的。

STEP 之外

這個密度函數是遞減的，所以它是「右偏」（skewed to the right）的：長長的尾部把平均值拉到中位數之外。這種現象很常見，但並非必然；大學的第一門概率課程會說明，對於單峯分佈（unimodal distribution），平均值、中位數和眾數（mode）之間有何關係，以及那些耳熟能詳的經驗法則在甚麼情況下會失效。像這樣的階梯函數，也是直方圖（histogram）的基本組件，同時是測度論（measure theory）中用來定義積分的簡單函數（simple function）的基本組件。想進一步了解連續分佈，可參閱 Grimmett and Welsh, *Probability: An Introduction*（概率論入門教材）。

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。