详细解答

STEP II 2010，第 12 题

概率与统计 • 平均值与中位数

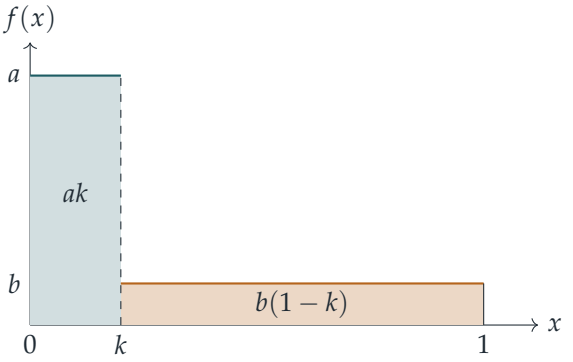
STEP II 2010，第 12 题

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/10-s2-q12

关键思路

这个密度函数（density function）是只有两级的阶梯函数（step function），所以每个概率都是一个长方形的面积，而唯一真正的条件就是总面积等于 1。这个条件把 k 用 a 和 b 表示出来；一旦消去 k，余下的全是代数运算。至于中位数（median），问题只在于那个较高的第一个长方形是否已经占了一半的概率。



解答

图中两个长方形的面积分别是 ak 和 $b(1-k)$，所以密度函数的积分等于 1，当且仅当

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

（由于 $a > b$，我们可以除以 $a-b$。）以下一直采用这个 k 值。

为何 $a > 1$ 且 $b < 1$。 方程 $ak + b(1-k) = 1$ 表示 1 是 a 和 b 的加权平均（weighted average），权重为 k 和 $1-k$，两者都严格介乎 0 与 1 之间。权重为正的加权平均必定严格介乎两个数值之间，所以 $b < 1 < a$。不用文字来说，就是：$1 - b = k(a-b) > 0$ 及 $a - 1 = (1-k)(a-b) > 0$。

第 (i) 部分

在每一段上对 $xf(x)$ 积分，

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

现在 $(a-b)k^2 = \dfrac{(1-b)^2}{a-b}$，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

累积分布函数（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 $[0,1]$ 上连续而且严格递增，所以中位数 M 是 $\mathrm{F}(M) = \frac{1}{2}$ 的唯一解。该用哪一条公式，取决于第一个长方形是否占了至少一半的概率。

> **想一想** 怎样把“第一个长方形的面积至少是 $\frac{1}{2}$”化为关于 a 和 b 的条件?

代入 k 的值，并消去正数分母 $a-b$：

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

情况 $a + b \ge 2ab$。这时 $\mathrm{F}(k) = ak \ge \frac{1}{2}$，所以中位数位于 $[0,k]$ 之内，并满足 $aM = \frac{1}{2}$：

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

情况 $a + b \le 2ab$。这时 $ak \le \frac{1}{2}$，所以中位数位于 $[k,1]$ 之内。从右端入手最为简便：M 右方的概率是 $b(1-M)$，而它必须等于 $\frac{1}{2}$。因此

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

当 $a + b = 2ab$ 时，两条公式都适用，而且结果一致，因为这时 $1 - \frac{1}{2b} = \frac{2ab-a}{2ab} = \frac{b}{2ab} = \frac{1}{2a}$。

第 (iii) 部分

我们在每个情况下计算 $\mathrm{E}(X)-M$，并证明它是正数。

情况 $a+b\geq 2ab$。通分，以 $2a(a-b)$ 为公分母，

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

情况 $a+b\leq 2ab$。通分，以 $2b(a-b)$ 为公分母，

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

想一想 为甚么这些数值是严格为正，而不仅仅是非负?

在两个情况下，a、b 和 $a-b$ 都是正数，而平方项严格为正，因为我们一开始已证明 $a>1$ 且 $b<1$。所以在任何情况下都有 $\mathrm{E}(X)-M>0$，即 $M<\mathrm{E}(X)$。

验算

- 取 $a=3$、$b=\frac{1}{2}$：这时 $k=\frac{1}{5}$，$a+b=\frac{7}{2}\geq 3=2ab$，而 $\mathrm{E}(X)=\frac{1-1+\frac{3}{2}}{5}=0.3$，$M=\frac{1}{6}$。两者之差 $0.1\overline{3}$ 与 $\frac{b(a-1)^2}{2a(a-b)}=\frac{2}{15}$ 吻合。
- 取 $a=\frac{5}{4}$、$b=\frac{9}{10}$：这次 $a+b=2.15\leq 2.25=2ab$，所以 $M=1-\frac{5}{9}=\frac{4}{9}\approx 0.444$，而 $\mathrm{E}(X)=\frac{13}{28}\approx 0.464$。
- 当 $a\to 1^+$（因而 $b\to 1^-$）时，密度函数趋向均匀分布（uniform distribution）的密度函数，两个差值都趋向 0：平均值（mean）与中位数同为 $\frac{1}{2}$，按对称性理应如此。

我们以 300 对随机选取的 (a,b) 作了数值检验：对 F 使用二分法（bisection）求得的中位数与分情况的公式吻合，以中点求积法（midpoint quadrature）求得的平均值与第 (i) 部分吻合，而 $\mathrm{E}(X)-M$ 亦等于上面已因式分解的表达式。

常见错误

- 忘记消去 k：答案须只用 a 和 b 表示。
- 把第二个中位数写成 $b(M-k)=\frac{1}{2}$ 的解，忘记了在 $[0,k)$ 上已经累积的概率 ak。
- 只得出 $\mathrm{E}(X)\geq M$ 的结论：严格不等式需要 $a\neq 1$ 和 $b\neq 1$，而这正是题目开首部分所提供的。

STEP 之外

这个密度函数是递减的，所以它是“右偏”（skewed to the right）的：长长的尾部把平均值拉到中位数之外。这种现象很常见，但并非必然；大学的第一门概率课程会说明，对于单峰分布（unimodal distribution），平均值、中位数和众数（mode）之间有何关系，以及那些耳熟能详的经验法则在甚么情况下会失效。像这样的阶梯函数，也是直方图（histogram）的基本组件，同时是测度论（measure theory）中用来定义积分的简单函数（simple function）的基本组件。想进一步了解连续分布，可参阅 Grimmett and Welsh, *Probability: An Introduction*（概率论入门教材）。

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。