解答例

STEP II 2010、第 12 問

確率・統計 ・ 平均値と中央値

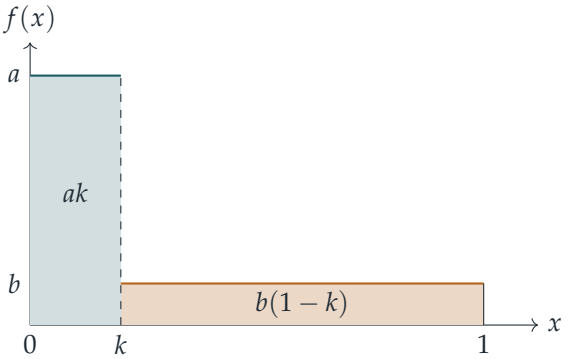
STEP II 2010、問 12

先に進む前に、まず問題を読んでおこう。問題は STEP Support Programme の問題データベースにある：

step.maths.org/questions/10-s2-q12

鍵となる考え方

密度関数（density function）は 2 段の階段関数（step function）なので、どの確率も長方形の面積であり、本質的な条件は全面積が 1 であることだけである。この条件によって k が a と b で表され、k を消去してしまえば、あとはすべて代数計算である。中央値（median）については、背の高い最初の長方形だけですでに確率の半分を占めているかどうか、という問題にすぎない。



解答

図の 2 つの長方形の面積は ak と $b(1-k)$ であるから、密度関数の積分が 1 になるのは

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

のときであり、かつそのときに限る（$a > b$ であるから $a - b$ で割ってよい）。以下ではこの k の値を用いる。

$a > 1$ かつ $b < 1$ である理由。 等式 $ak + b(1-k) = 1$ は、1 が a と b の重み k と $1-k$ による加重平均（weighted average）であることを表しており、どちらの重みも 0 と 1 の間に厳密に含まれる。正の重みによる加重平均は 2 つの値の厳密に間にあるので、$b < 1 < a$ である。言葉を使わずに表せば、$1 - b = k(a-b) > 0$ かつ $a - 1 = (1-k)(a-b) > 0$ である。

小問 (i)

各区間で $xf(x)$ を積分すると、

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

ここで $(a-b)k^2 = \dfrac{(1-b)^2}{a-b}$ であるから、

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

となり、題意が示された。

小問 (ii)

累積分布関数（cumulative distribution function）は

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

である。これは $[0,1]$ 上で連続かつ狭義単調増加であるから、中央値 M は $\mathrm{F}(M) = \frac{1}{2}$ のただ 1 つの解である。どちらの式が当てはまるかは、最初の長方形が確率の半分以上を占めるかどうかによって決まる。

> **考えてみよう** 「最初の長方形の面積が $\frac{1}{2}$ 以上である」ことを、a と b に関する条件に書き換えるにはどうすればよいか。

k を代入し、正の分母 $a-b$ を払うと、

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

$a + b \ge 2ab$ の場合。 このとき $\mathrm{F}(k) = ak \ge \frac{1}{2}$ であるから、中央値は $[0,k]$ にあり、$aM = \frac{1}{2}$ を満たす。よって

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

$a+b \le 2ab$ **の場合。** このとき $ak \le \frac{1}{2}$ であるから、中央値は $[k,1]$ にある。右端から考えるのが最も簡単である。M より右側の確率は $b(1-M)$ であり、これが $\frac{1}{2}$ に等しくなければならない。したがって

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

$a+b = 2ab$ のときは両方の式が当てはまり、しかも一致する。実際、このとき $1-\frac{1}{2b} = \frac{2ab-a}{2ab} = \frac{b}{2ab} = \frac{1}{2a}$ である。

小問 (iii)

それぞれの場合について $\mathrm{E}(X) - M$ を計算し、それが正であることを示す。

$a+b \ge 2ab$ **の場合。** 共通分母 $2a(a-b)$ で通分すると、

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

$a+b \le 2ab$ **の場合。** 共通分母 $2b(a-b)$ で通分すると、

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

考えてみよう これらが単に非負であるだけでなく、厳密に正であるのはなぜか。

どちらの場合も a、b、$a-b$ は正であり、最初に $a>1$ かつ $b<1$ を示したので、平方の部分も厳密に正である。よって、どの場合でも $\mathrm{E}(X) - M > 0$ であり、$M < \mathrm{E}(X)$ が成り立つ。

検算

- $a=3$、$b=\frac{1}{2}$ とする。このとき $k=\frac{1}{5}$、$a+b=\frac{7}{2} \ge 3 = 2ab$ であり、$\mathrm{E}(X) = \frac{1-1+\frac{3}{2}}{5} = 0.3$、一方 $M=\frac{1}{6}$ である。差 $0.1\overline{3}$ は $\frac{b(a-1)^2}{2a(a-b)} = \frac{2}{15}$ と一致する。
- $a=\frac{5}{4}$、$b=\frac{9}{10}$ とする。今度は $a+b=2.15 \le 2.25 = 2ab$ であるから、$M = 1-\frac{5}{9} = \frac{4}{9} \approx 0.444$、また $\mathrm{E}(X) = \frac{13}{28} \approx 0.464$ である。
- $a \to 1^+$（したがって $b \to 1^-$）のとき、密度関数は一様分布の密度関数に近づき、どちらの差も 0 に近づく。対称性から当然そうあるべきように、平均（mean）と中央値は $\frac{1}{2}$ で一致する。

300 組の無作為な (a, b) について、F に二分法（bisection）を適用して求めた中央値が場合分けの式と一致すること、中点則（midpoint quadrature）による平均が小問 (i) の結果と一致すること、そして $\mathrm{E}(X) - M$ が上の因数分解された式に等しいことを数値的に確認した。

よくある間違い

- k を消去し忘れる。答えは a と b だけで表すことが求められている。
- 2 つ目の中央値を $b(M-k) = \frac{1}{2}$ の解として書いてしまい、$[0, k)$ ですでに蓄積された確率 ak を忘れる。
- $\mathrm{E}(X) \geq M$ しか結論しない。厳密な不等式には $a \neq 1$ かつ $b \neq 1$ が必要であり、それこそが冒頭の部分で示されたことである。

STEP のその先へ

この密度関数は減少関数なので「右に歪んでいる（skewed to the right）」。長い裾が平均を中央値より先へ引っ張るのである。このパターンはよく見られるが普遍的ではなく、確率論の最初の講義では、単峰（unimodal）分布について平均、中央値、最頻値（mode）がどのような関係にあるか、そしてよく知られた経験則がどこで成り立たなくなるかを学ぶ。このような階段関数は、ヒストグラム（histogram）の構成要素でもあり、また測度論（measure theory）で積分を定義する際に用いられる単関数（simple function）の構成要素でもある。連続分布についてさらに学ぶには、Grimmett and Welsh, *Probability: An Introduction*（確率論入門）を参照されたい。

Beyond STEP 解答集：https://beyondstep.keithmcnulty.org/ja/。解答本文は CC0 1.0 である。問題そのものは © Cambridge University Press & Assessment に帰属し、ここには掲載していない。英語版から機械翻訳したものであり、まだ校閲を経ていない。