解答例

STEP II 2007、第 14 問

確率・統計 ・ 区分的な確率密度

STEP II 2007、問 14

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/07-s2-q14

鍵となる考え方

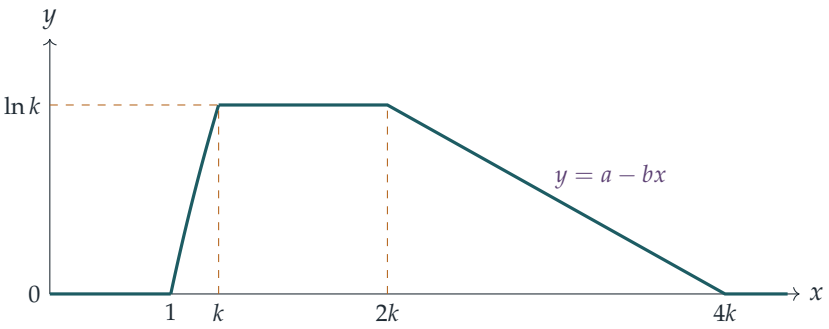
連続（continuous）という言葉が大きな役割を果たしている。区間ごとに定義された確率密度関数（probability density function）は、すべての境目でつながっていなければならず、そこからただちに a と b についての方程式が 2 本得られる。残る条件、すなわちグラフの下の面積の総和が 1 であることから、k が決まる。中央値（median）を求めるには、面積を区間ごとに足し合わせていき、2 分の 1 を超えるところを探す。

解答

以下では常に $k > 1$ とする。そうでなければ区間 $1 \leq x \leq k$ は空になってしまうし、いずれにせよ、密度関数が負にならず、かつ $x = 1$ より先で恒等的に 0 にならないためには $\ln k > 0$ が必要である。

小問 (i)

$[1, k]$ 上ではグラフは通常の対数曲線であり、$x = 1$ で 0 から始まり（したがって値が 0 の部分と連続につながる）、単調増加で上に凸（concave）である。$x = k$ で高さ $\ln k$ に達し、$x = 2k$ まで $\ln k$ の高さで一定に保たれ、その後 $x = 4k$ で 0 になるまで直線的に減少し（連続性によってまさにこうなることを以下で示す）、それ以降は 0 である。図には小問 (ii) で求める値 $k = \mathrm{e}^{1/3} \approx 1.40$ を用いており、したがって平坦部の高さは $\frac{1}{3}$ である。



小問 (ii)

連続性。 $x = k$ ではどちらの式も $\ln k$ を与えるので、新しい情報は得られない。$x = 2k$ と $x = 4k$ では、連続性から

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

が要求される。辺々引くと $2bk = \ln k$ となるので、

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

全確率。 面積を 3 つの部分に分ける。部分積分（integration by parts）により（$u = \ln x$, $\mathrm{d}v = \mathrm{d}x$ とする）、

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

平坦部は幅 k、高さ $\ln k$ の長方形であり、面積は $k\ln k$ である。最後の部分は底辺 $4k - 2k = 2k$、高さ $\ln k$ の三角形であり、面積は $\frac{1}{2} \cdot 2k \cdot \ln k = k\ln k$ である。よって

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

であり、これを整理すると

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

となる。$k > 1$ より $k = 0$ はありえないので、$\ln k = \frac{1}{3}$ であり、

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

$b > 0$ であるから、最後の部分は図のとおり確かに右下がりになっていることに注意する。

小問 (iii)

確率を累積していくことで、中央値 m がどの部分に含まれるかを調べる。$\ln k = \frac{1}{3}$ を用いると、

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

であり、平坦部が $k\ln k = \frac{k}{3}$ を加えるので、

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

考えてみよう 中央値はグラフのどの部分にあるか？

$0.070 < \frac{1}{2} < 0.535$ であるから、中央値は平坦部、すなわち $k \le m \le 2k$ にある。そこでは密度関数は $\frac{1}{3}$ で一定なので、

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

両辺に 3 を掛けると $3 - 2k + m - k = \frac{3}{2}$ となるので、

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

これは、あるべきとおり、確かに $k \approx 1.40$ と $2k \approx 2.79$ の間にある。

検算

- $k = \mathrm{e}^{1/3}$、$a = \frac{2}{3}$、$b = 1/(6\mathrm{e}^{1/3})$ として密度関数を数値積分（中点則）すると、面積の総和は 1.0000000 となり、また $\int_1^{2.68684} f(x)\,\mathrm{d}x = 0.5000000$ となる。
- 中央値の妥当性の確認：$2k$ より左にある確率は 2 分の 1 をわずかに超える（約 0.535）ので、中央値は $2k \approx 2.79$ の少し左にあるはずである。実際 $m \approx 2.69$ である。

よくある間違い

- $\int_1^k \ln x\,\mathrm{d}x$ の下端を忘れること。$x = 1$ における $-[x\ln x - x]$ から生じる $+1$ こそが、k についての方程式をきれいに因数分解できる形にしている。
- 台 $[1, 4k]$ の「真ん中」にあるからという理由で、中央値が右下がりの部分にあると思い込むこと。必ず先に面積を累積すること。

STEP のその先へ

ここでは密度関数を区間ごとの部分から組み立て、定数を連続性と正規化（normalisation）によって決定した。大学では、**累積分布関数**（cumulative distribution function）$F(x) = \mathrm{P}(X \le x)$ を基本的な対象として学ぶ。これは離散型・連続型を問わず、あらゆる確率変数（random variable）に対して存在し、中央値のような量は単に $F(m) = \frac{1}{2}$ の解として与えられる。Grimmett と Welsh の *Probability: An Introduction*（確率論入門）は、この観点をわかりやすく展開している。

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。