解答例

STEP I 2014、第 13 問

確率・統計 ・ 三角分布

STEP I 2014、問 13

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：

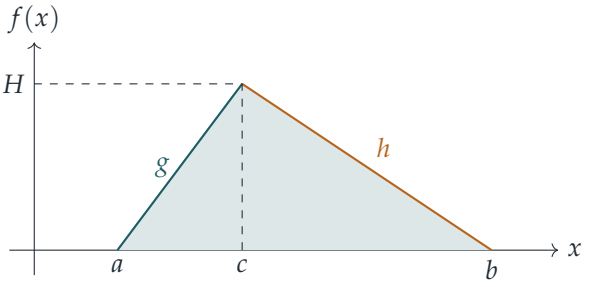
step.maths.org/questions/14-s1-q13

鍵となる考え方

確率密度関数（density）のグラフは、底辺が a から b まで、頂点（apex）が c の真上にある三角形であり、その面積は 1 でなければならない。これで頂点の高さが決まり、あとは各線形（linear）部分が 2 点によって定まる。平均（mean）と中央値（median）については、各線形部分が 0 から始まるように変数をずらすと積分がすっきりする。さらに中央値については、まずそれが c のどちら側にあるかを考えればよい。

解答

以下では常に $a < c < b$ とし、どちらの部分も実際に存在するものとする（問題の公式が意味をもつためには、$c - a$ と $b - c$ が 0 でないことが必要である）。



2 つの線形部分

頂点の高さを $H = g(c) = h(c)$ とおく。f のグラフの下の領域は底辺 $b-a$、高さ H の三角形であり、全確率は 1 であるから、

$$\tfrac{1}{2}(b-a)H = 1, \qquad H = \frac{2}{b-a}.$$

g は線形で $g(a) = 0$ であるから、ある定数 k を用いて $g(x) = k(x-a)$ と書ける。$g(c) = H$ より $k(c-a) = \dfrac{2}{b-a}$ である。よって

$$g(x) = \frac{2(x-a)}{(b-a)(c-a)},$$

となり、求める式が得られた。同様に $h(x) = m(b-x)$ かつ $h(c) = H$ であるから、

$$h(x) = \frac{2(b-x)}{(b-a)(b-c)}.$$

$b-a$、$c-a$、$b-c$ はすべて正であるから、要求どおり g は増加関数、h は減少関数である。

小問 (i)

$\mathrm{E}(X) = I_1 + I_2$ と書く。ただし $I_1 = \int_a^c x\,g(x)\,\mathrm{d}x$、$I_2 = \int_c^b x\,h(x)\,\mathrm{d}x$ である。

I_1 において $x = a + t$ と置換すると、t は 0 から $c-a$ まで動く：

$$\begin{aligned} I_1 &= \frac{2}{(b-a)(c-a)} \int_0^{c-a} (a+t)\,t\,\mathrm{d}t = \frac{2}{(b-a)(c-a)} \left[\frac{a(c-a)^2}{2} + \frac{(c-a)^3}{3}\right] \\ &= \frac{(c-a)(3a + 2(c-a))}{3(b-a)} = \frac{(c-a)(a+2c)}{3(b-a)}. \end{aligned}$$

I_2 において $x = b - s$ と置換すると、s は 0 から $b-c$ まで動く：

$$\begin{aligned} I_2 &= \frac{2}{(b-a)(b-c)} \int_0^{b-c} (b-s)\,s\,\mathrm{d}s = \frac{2}{(b-a)(b-c)} \left[\frac{b(b-c)^2}{2} - \frac{(b-c)^3}{3}\right] \\ &= \frac{(b-c)(3b - 2(b-c))}{3(b-a)} = \frac{(b-c)(b+2c)}{3(b-a)}. \end{aligned}$$

これらを足すと、分子は

$$\begin{aligned} (c-a)(a+2c) + (b-c)(b+2c) &= (ac + 2c^2 - a^2 - 2ac) + (b^2 + 2bc - bc - 2c^2) \\ &= b^2 - a^2 + bc - ac \\ &= (b-a)(b+a+c). \end{aligned}$$

$3(b-a)$ で割ると、

$$\mathrm{E}(X) = \tfrac{1}{3}(a+b+c),$$

となり、求める結果が得られた。（これは三角形の重心（centroid）の x 座標であり、偶然ではない。平均とは、密度関数の下の領域の釣り合いの点なのである。）

小問 (ii)

まず、頂点より左側にどれだけの確率があるかを求める。これは左側の三角形の面積である：

$$\mathrm{P}(X \le c) = \tfrac{1}{2}(c-a)H = \frac{c-a}{b-a}.$$

考えてみよう 中央値が c の左にあるか右にあるかは、どうすれば分かるだろうか？

中央値 M は $\mathrm{P}(X \le M) = \frac{1}{2}$ を満たす。$\mathrm{P}(X \le c) \ge \frac{1}{2}$ ならば中央値は c 上またはその左にあり、そうでなければ右にある。ここで $\frac{c-a}{b-a} \ge \frac{1}{2}$ となるのは $c \ge \frac{1}{2}(a+b)$ のとき、すなわち頂点が $[a,b]$ の中点上またはその右にあるときに限る。したがって、場合分けは c と $\frac{1}{2}(a+b)$ の比較によって決まる。

$c \ge \frac{1}{2}(a+b)$ の場合。 このとき $a < M \le c$ であり、$a \le m \le c$ に対して、a から m までの小さな三角形の面積は

$$\mathrm{P}(X \le m) = \int_a^m g(x)\,\mathrm{d}x = \frac{(m-a)^2}{(b-a)(c-a)}.$$

これを $\frac{1}{2}$ に等しいとおき、正の根をとると（$M > a$ であるから）、

$$M = a + \sqrt{\tfrac{1}{2}(b-a)(c-a)}.$$

$c < \frac{1}{2}(a+b)$ の場合。 このとき中央値は c の右にあり、右端から考える方が簡単である。$c \le m \le b$ に対して、

$$\mathrm{P}(X \ge m) = \int_m^b h(x)\,\mathrm{d}x = \frac{(b-m)^2}{(b-a)(b-c)}.$$

これを $\frac{1}{2}$ に等しいとおき、$M < b$ を用いると、

$$M = b - \sqrt{\tfrac{1}{2}(b-a)(b-c)}.$$

$c = \frac{1}{2}(a+b)$ のとき分布は対称（symmetric）であり、どちらの公式も $M = \frac{1}{2}(a+b)$ を与える（たとえば 1 つ目の公式からは $a + \sqrt{\frac{1}{4}(b-a)^2} = a + \frac{1}{2}(b-a)$ が得られる）。したがって 2 つの場合はつながっている。

検算

- $a = 0$、$b = 4$、$c = 1$ のとき：$\mathrm{P}(X \le c) = \frac{1}{4} < \frac{1}{2}$ であるから 2 つ目の場合にあたり、$M = 4 - \sqrt{6} \approx 1.5505$ である。平均は $\frac{5}{3}$ である。コンピュータで 200 000 個の値をシミュレーションしたところ、標本中央値は 1.5505、標本平均は 1.668 であった。

- $a = 0$、$b = 4$、$c = 3$ のとき：1 つ目の場合にあたり、$M = \sqrt{6} \approx 2.449$ である。これは当然そうあるべきとおり、前の例の鏡像になっている。
- 中央値は常に最頻値（mode）c と中点 $\frac{1}{2}(a+b)$ の間にある。たとえば 1 つ目の場合には $M \leq c$ であり、また $c \geq \frac{1}{2}(a+b)$ のとき $(b-a)(c-a) \geq \frac{1}{2}(b-a)^2$ であるから $M \geq \frac{1}{2}(a+b)$ となる。

よくある間違い

- 中央値を求める際に、密度関数の誤った部分を使ってしまう。まず中央値が c のどちら側にあるかを確かめること。
- 負の平方根をとってしまい、M が $[a,b]$ の外に出てしまう。
- 平均の積分を、変数をずらさずに x のまま直接展開する。それでも計算はできるが、式変形がずっと重くなり、誤りが入り込みやすい。

STEP のその先へ

三角分布（triangular distribution）は、c が中点のとき、2 つの独立（independent）な一様確率変数（uniform random variable）の和の分布になっている。また、最小値・最大値・最も起こりやすい値だけが分かっている場合の大まかなモデルとして（たとえばプロジェクト計画において）広く用いられている。独立な確率変数の和の分布を計算すると、畳み込み（convolution）に行き着く。これは大学 1 年次の確率論における中心的な道具である。これについて読むには、Grimmett and Welsh, *Probability: An Introduction*（確率論入門）がよい。

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。