详细解答

STEP I 2014，第 13 题

概率与统计 • 三角分布

STEP I 2014，第 13 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：

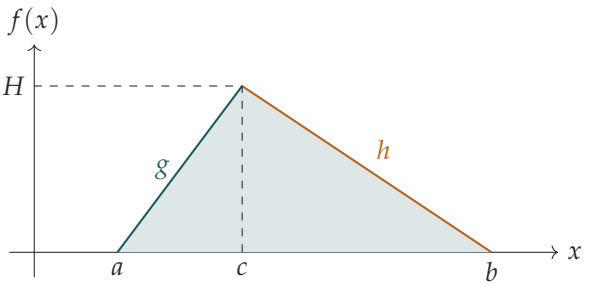
step.maths.org/questions/14-s1-q13

关键思路

密度函数（density function）的图像是一个三角形，底边由 a 延伸至 b，顶点（apex）位于 c 的正上方，而三角形的面积必须等于 1。这就确定了顶点的高度，之后每段线性部分都由两点完全确定。计算平均值（mean）和中位数（median）时，先平移变量，使每段线性部分都从零开始，积分就会保持简洁；至于中位数，首要问题只是它落在 c 的哪一边。

解答

全文假设 $a < c < b$，使两段都是真正存在的（题目的公式要有意义，$c - a$ 与 $b - c$ 必须不等于零）。



两段线性部分

设 $H = g(c) = h(c)$ 为顶点的高度。f 的图像下方的区域是一个底为 $b-a$、高为 H 的三角形，而总概率为 1，所以

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

由于 g 是线性函数且 $g(a) = 0$，故对某常数 k 有 $g(x) = k(x-a)$，而 $g(c) = H$ 给出 $k(c-a) = \dfrac{2}{b-a}$。因此

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

得证。同理，$h(x) = m(b-x)$ 且 $h(c) = H$，所以

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

由于 $b-a$、$c-a$ 和 $b-c$ 全部为正，g 是递增的而 h 是递减的，符合要求。

第 (i) 部分

记 $\mathrm{E}(X) = I_1 + I_2$，其中 $I_1 = \int_a^c x\,g(x)\,\mathrm{d}x$，$I_2 = \int_c^b x\,h(x)\,\mathrm{d}x$。

在 I_1 中代入 $x = a + t$，于是 t 由 0 变到 $c-a$：

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

在 I_2 中代入 $x = b - s$，于是 s 由 0 变到 $b-c$：

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

两者相加，分子为

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

除以 $3(b-a)$，得

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

得证。（这正是三角形形心（centroid）的 x 坐标，而这并非巧合：平均值就是密度函数下方区域的平衡点。）

第 (ii) 部分

首先求出顶点左方有多少概率。这就是左边那个三角形的面积：

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

想一想 我们怎样知道中位数在 c 的左边还是右边?

中位数 M 满足 $\mathrm{P}(X \le M) = \frac{1}{2}$。若 $\mathrm{P}(X \le c) \ge \frac{1}{2}$，中位数就在 c 处或其左方；否则就在其右方。而 $\frac{c-a}{b-a} \ge \frac{1}{2}$ 当且仅当 $c \ge \frac{1}{2}(a+b)$，即顶点位于 $[a,b]$ 的中点处或其右方。所以分情况的依据，就是比较 c 与 $\frac{1}{2}(a+b)$。

情况 $c \ge \frac{1}{2}(a+b)$。此时 $a < M \le c$，而对于 $a \le m \le c$，由 a 到 m 的小三角形面积为

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

令它等于 $\frac{1}{2}$，并取正根（因为 $M > a$），得

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

情况 $c < \frac{1}{2}(a+b)$。这时中位数在 c 的右方，从右端入手会比较容易。对于 $c \le m \le b$,

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

令它等于 $\frac{1}{2}$，并利用 $M < b$，得

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

当 $c = \frac{1}{2}(a+b)$ 时，分布是对称的，两条公式都给出 $M = \frac{1}{2}(a+b)$（例如第一条给出 $a + \sqrt{\frac{1}{4}(b-a)^2} = a + \frac{1}{2}(b-a)$），所以两种情况能够衔接起来。

验算

- 取 $a=0$、$b=4$、$c=1$：$\mathrm{P}(X \le c) = \frac{1}{4} < \frac{1}{2}$，所以属于第二种情况，$M = 4 - \sqrt{6} \approx 1.5505$。平均值为 $\frac{5}{3}$。用电脑模拟 200 000 个数值，得到样本中位数 1.5505 及样本平均值 1.668。
- 取 $a=0$、$b=4$、$c=3$：适用第一种情况，$M = \sqrt{6} \approx 2.449$，正好是上一个例子的镜像，理应如此。
- 中位数总是介乎众数（mode）c 与中点 $\frac{1}{2}(a+b)$ 之间。例如在第一种情况中 $M \le c$，而由于当 $c \ge \frac{1}{2}(a+b)$ 时 $(b-a)(c-a) \ge \frac{1}{2}(b-a)^2$，所以 $M \ge \frac{1}{2}(a+b)$。

常见错误

- 求中位数时用错了密度函数的那一段：应先确定中位数落在 c 的哪一边。
- 取了负的平方根，令 M 落在 $[a, b]$ 之外。
- 不作平移，直接以 x 展开平均值的积分：这样做也可以，但代数运算繁重得多，容易出错。

STEP 之外

当 c 是中点时，三角分布（triangular distribution）就是两个独立均匀随机变量（uniform random variable）之和的分布；而在只知道最小值、最大值和最可能值的情况下（例如项目规划），三角分布亦广泛用作粗略的模型。计算独立随机变量之和的分布会引出卷积（convolution），这是大学一年级概率论的核心工具。Grimmett 与 Welsh 的 *Probability: An Introduction*（《概率导论》）是阅读这方面内容的好起点。

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。