詳細解答

STEP I 2014，第 13 題

概率與統計 • 三角分佈

STEP I 2014，第 13 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

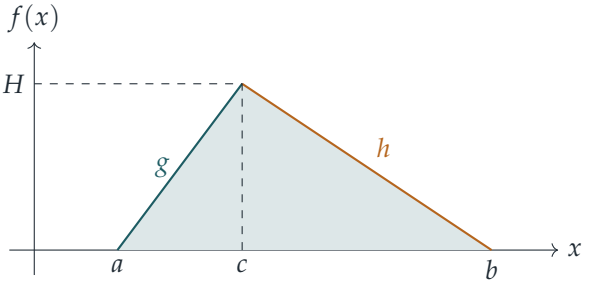
step.maths.org/questions/14-s1-q13

關鍵思路

密度函數（density function）的圖像是一個三角形，底邊由 a 延伸至 b，頂點（apex）位於 c 的正上方，而三角形的面積必須等於 1。這就確定了頂點的高度，之後每段線性部分都由兩點完全確定。計算平均值（mean）和中位數（median）時，先平移變量，使每段線性部分都從零開始，積分就會保持簡潔；至於中位數，首要問題只是它落在 c 的哪一邊。

解答

全文假設 $a < c < b$，使兩段都是真正存在的（題目的公式要有意義，$c - a$ 與 $b - c$ 必須不等於零）。



兩段線性部分

設 $H = g(c) = h(c)$ 為頂點的高度。f 的圖像下方的區域是一個底為 $b-a$、高為 H 的三角形，而總概率為 1，所以

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

由於 g 是線性函數且 $g(a) = 0$，故對某常數 k 有 $g(x) = k(x-a)$，而 $g(c) = H$ 給出 $k(c-a) = \dfrac{2}{b-a}$。因此

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

得證。同理，$h(x) = m(b-x)$ 且 $h(c) = H$，所以

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

由於 $b-a$、$c-a$ 和 $b-c$ 全部為正，g 是遞增的而 h 是遞減的，符合要求。

第 (i) 部分

記 $\mathrm{E}(X) = I_1 + I_2$，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 $x = a + t$，於是 t 由 0 變到 $c-a$：

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

在 I_2 中代入 $x = b - s$，於是 s 由 0 變到 $b-c$：

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

兩者相加，分子為

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

除以 $3(b-a)$，得

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

得證。（這正是三角形形心（centroid）的 x 坐標，而這並非巧合：平均值就是密度函數下方區域的平衡點。）

第 (ii) 部分

首先求出頂點左方有多少概率。這就是左邊那個三角形的面積：

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

想一想 我們怎樣知道中位數在 c 的左邊還是右邊？

中位數 M 滿足 $\mathrm{P}(X \le M) = \frac{1}{2}$。若 $\mathrm{P}(X \le c) \ge \frac{1}{2}$，中位數就在 c 處或其左方；否則就在其右方。而 $\frac{c-a}{b-a} \ge \frac{1}{2}$ 當且僅當 $c \ge \frac{1}{2}(a+b)$，即頂點位於 $[a,b]$ 的中點處或其右方。所以分情況的依據，就是比較 c 與 $\frac{1}{2}(a+b)$。

情況 $c \ge \frac{1}{2}(a+b)$。此時 $a < M \le c$，而對於 $a \le m \le c$，由 a 到 m 的小三角形面積為

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

令它等於 $\frac{1}{2}$，並取正根（因為 $M > a$），得

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

情況 $c < \frac{1}{2}(a+b)$。這時中位數在 c 的右方，從右端入手會比較容易。對於 $c \le m \le b$，

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

令它等於 $\frac{1}{2}$，並利用 $M < b$，得

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

當 $c = \frac{1}{2}(a+b)$ 時，分佈是對稱的，兩條公式都給出 $M = \frac{1}{2}(a+b)$（例如第一條給出 $a + \sqrt{\frac{1}{4}(b-a)^2} = a + \frac{1}{2}(b-a)$），所以兩種情況能夠銜接起來。

驗算

- 取 $a=0$、$b=4$、$c=1$：$\mathrm{P}(X \le c) = \frac{1}{4} < \frac{1}{2}$，所以屬於第二種情況，$M = 4 - \sqrt{6} \approx 1.5505$。平均值為 $\frac{5}{3}$。用電腦模擬 200 000 個數值，得到樣本中位數 1.5505 及樣本平均值 1.668。
- 取 $a=0$、$b=4$、$c=3$：適用第一種情況，$M = \sqrt{6} \approx 2.449$，正好是上一個例子的鏡像，理應如此。
- 中位數總是介乎眾數（mode）c 與中點 $\frac{1}{2}(a+b)$ 之間。例如在第一種情況中 $M \le c$，而由於當 $c \ge \frac{1}{2}(a+b)$ 時 $(b-a)(c-a) \ge \frac{1}{2}(b-a)^2$，所以 $M \ge \frac{1}{2}(a+b)$。

常見錯誤

- 求中位數時用錯了密度函數的那一段：應先確定中位數落在 c 的哪一邊。
- 取了負的平方根，令 M 落在 $[a, b]$ 之外。
- 不作平移，直接以 x 展開平均值的積分：這樣做也可以，但代數運算繁重得多，容易出錯。

STEP 之外

當 c 是中點時，三角分佈（triangular distribution）就是兩個獨立均勻隨機變量（uniform random variable）之和的分佈；而在只知道最小值、最大值和最可能值的情況下（例如項目規劃），三角分佈亦廣泛用作粗略的模型。計算獨立隨機變量之和的分佈會引出卷積（convolution），這是大學一年級概率論的核心工具。Grimmett 與 Welsh 的 *Probability: An Introduction*（《概率導論》）是閱讀這方面內容的好起點。

Beyond STEP 詳細解答：https://beyondstep.keithmcnulty.org/zh-hant/。解答文字採用 CC0 1.0；題目本身版權屬 Cambridge University Press & Assessment 所有（©），本文件不予轉載。由機器從英文翻譯而成；尚未經審閱。