詳細解答

STEP I 2005，第 13 題

概率與統計 • 對稱分佈

> **STEP I 2005，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/05-s1-q13

關鍵思路

題目除了給出兩個累積概率（cumulative probability）和對稱性（symmetry）之外，對分佈的形狀隻字未提，所以一切都要由這些資料推導出來。關於 μ 對稱，還附帶送來兩個事實：μ 的兩側各佔一半概率，而且 $\mu - k\sigma$ 以下的尾部與 $\mu + k\sigma$ 以上的尾部概率相同。第 (ii) 部分的訣竅，是留意到題中提及的每一個容量（495、500、505 ml），對其中一種牛奶而言，都正好落在 μ、$\mu \pm \frac{1}{2}\sigma$ 或 $\mu + \sigma$ 的位置。

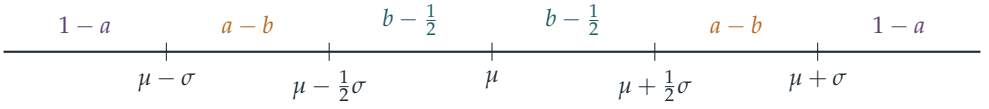
解答

全題把 X 視為連續（continuous）隨機變量（它描述的是容量），因此單一數值的概率為零，寫 $<$ 還是 $\leq$ 並無分別。

由關於 μ 的對稱性，可得 $\mathrm{P}(X \leq \mu) = \frac{1}{2}$，並且對任意 k 都有 $\mathrm{P}(X \leq \mu - k\sigma) = \mathrm{P}(X \geq \mu + k\sigma)$。特別地，

$$\mathrm{P}(X \leq \mu - \tfrac{1}{2}\sigma) = 1 - b, \qquad \mathrm{P}(X \leq \mu - \sigma) = 1 - a.$$

不妨以 σ 為單位，從 μ 起量度，把每一段區間的概率記下來：



例如 $\mathrm{P}(\mu < X \leq \mu + \frac{1}{2}\sigma) = b - \frac{1}{2}$，而 $\mathrm{P}(\mu + \frac{1}{2}\sigma < X \leq \mu + \sigma) = a - b$；左邊各段的概率則由對稱性得出。

第 (i) 部分

把累積概率相減，

$$\mathrm{P}\big(\mu - \tfrac{1}{2}\sigma \le X \le \mu + \sigma\big) = \mathrm{P}(X \le \mu + \sigma) - \mathrm{P}\big(X < \mu - \tfrac{1}{2}\sigma\big) = a - (1 - b) = a + b - 1.$$

至於條件概率（conditional probability），分子所對應的事件是 $\{\mu - \frac{1}{2}\sigma \le X \le \mu + \frac{1}{2}\sigma\}$，其概率為 $b - (1 - b) = 2b - 1$，而 $\mathrm{P}(X \ge \mu - \frac{1}{2}\sigma) = 1 - (1 - b) = b$。因此

$$\mathrm{P}\big(X \le \mu + \tfrac{1}{2}\sigma \mid X \ge \mu - \tfrac{1}{2}\sigma\big) = \frac{2b - 1}{b}.$$

第 (ii)(a) 部分

設 S 為該盒牛奶是脫脂奶的事件，F 為該盒牛奶是全脂奶的事件，則 $\mathrm{P}(S) = 0.6$ 及 $\mathrm{P}(F) = 0.4$。設 V 為以 ml 計的容量。

> **想一想** 以 σ 為單位，500 ml 和 505 ml 在每種牛奶的分佈中各落在甚麼位置？

對脫脂奶（$\mu = 500$，$\sigma = 10$）而言，500 是 μ，505 是 $\mu + \frac{1}{2}\sigma$。對全脂奶（$\mu = 495$，$\sigma = 10$）而言，500 是 $\mu + \frac{1}{2}\sigma$，505 是 $\mu + \sigma$。從上圖各段讀出：

$$\mathrm{P}(V < 505 \mid S) = b, \qquad \mathrm{P}(500 < V < 505 \mid S) = b - \tfrac{1}{2},$$
$$\mathrm{P}(V < 505 \mid F) = a, \qquad \mathrm{P}(500 < V < 505 \mid F) = a - b.$$

根據全概率公式（law of total probability），

$$\mathrm{P}(V < 505) = 0.6b + 0.4a,$$
$$\mathrm{P}(500 < V < 505) = 0.6\big(b - \tfrac{1}{2}\big) + 0.4(a - b) = 0.4a + 0.2b - 0.3.$$

「多於 500 ml 而少於 505 ml」這個事件正是我們所需的交集，所以

$$\mathrm{P}(V > 500 \mid V < 505) = \frac{0.4a + 0.2b - 0.3}{0.4a + 0.6b} = \frac{4a + 2b - 3}{4a + 6b}.$$

第 (ii)(b) 部分

隨機抽取的牛奶盒數量非常大時，觀察所得的比例可用來估計相應的概率。第一項資料表示 $\mathrm{P}(V \le 505) = 0.7$，按 (a) 部分的計算，即

$$0.6b + 0.4a = 0.7, \quad \text{即} \quad 4a + 6b = 7. \tag{1}$$

> **想一想** 對每種牛奶而言，$\mathrm{P}(V \ge 495)$ 是多少？

對脫脂奶而言，495 是 $\mu - \frac{1}{2}\sigma$，所以 $\mathrm{P}(V \geq 495 \mid S) = b$（這正是第 (i) 部分的分母）。對全脂奶而言，495 是 μ，所以由對稱性得 $\mathrm{P}(V \geq 495 \mid F) = \frac{1}{2}$。根據貝葉斯定理（Bayes' theorem），

$$\mathrm{P}(F \mid V \geq 495) = \frac{0.4 \times \frac{1}{2}}{0.6b + 0.4 \times \frac{1}{2}} = \frac{0.2}{0.6b + 0.2}.$$

令此式等於 $\frac{1}{3}$，得 $0.6b + 0.2 = 0.6$，所以 $b = \frac{2}{3}$。於是由 (1) 得 $4a = 7 - 4 = 3$，所以

$$a \approx \tfrac{3}{4}, \qquad b \approx \tfrac{2}{3}.$$

這些只是估計值，因為觀察所得的比例只是真實概率的近似。

驗算

- 估計值合情合理：$\frac{1}{2} < b < a < 1$，這是必然的，因為 $\mu < \mu + \frac{1}{2}\sigma < \mu + \sigma$。
- 代回檢驗：$0.6 \cdot \frac{2}{3} + 0.4 \cdot \frac{3}{4} = 0.4 + 0.3 = 0.7$，以及 $0.2/(0.4 + 0.2) = \frac{1}{3}$。
- 若分佈為正態分佈（Normal distribution），則 $a \approx 0.841$ 及 $b \approx 0.691$；以這些數值計算，第 (i) 部分得出 0.533 和 0.553。以（非正態的）拉普拉斯分佈（Laplace distribution）作模擬，樣本數目超過 10^5，結果與第 (i) 及 (ii)(a) 部分的所有公式吻合；例如 (ii)(a) 的模擬結果為 0.2518，而公式給出 0.2515。

常見錯誤

- 把 $\mathrm{P}(X \leq \mu - \frac{1}{2}\sigma) = b$ 當作正確，而其實應是 $1 - b$：對稱性反射的是*尾部*，而不是累積概率。
- 忘記兩種牛奶的平均值不同，所以同一個容量（例如 500 ml）對兩種牛奶而言，與平均值相距的標準差（standard deviation）數目並不相同。
- 對兩種牛奶都用 $\mathrm{P}(V \geq 495) = b$；對全脂奶而言，495 ml 正是平均值。

STEP 之外

這道題的要點在於：即使對分佈一無所知，單憑對稱性和全概率公式也可以走得很遠；這裏由兩個總體混合而成的情況，是一個簡單的*混合模型*（mixture model），而第 (ii)(b) 部分則是根據數據估計其參數的入門版本。另一個同樣不依賴分佈形狀的相關結果是切比雪夫（Chebyshev）不等式，它只用方差（variance）便能給出 $\mathrm{P}(|X - \mu| \geq k\sigma)$ 的上界。可參閱 Blitzstein and Hwang, *Introduction to Probability*（概率論入門），或 Grimmett and Welsh, *Probability: An Introduction*（概率論導引）。

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。