詳細解答

STEP III 2010，第 1 題

純數學 • 平均值與方差

> **STEP III 2010，第 1 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/10-s3-q1

關鍵思路

A 和 B 是首 n 個數的平均值（mean）和方差（variance）；C 和 D 則是再加入一個數之後的平均值和方差。方差最容易以「平方的平均值減去平均值的平方」這個形式處理，而這正是第 (ii) 部分給我們的結果。有了這個形式，D 便成為 x_{n+1} 的二次式（quadratic），而兩條不等式（inequality）都是關於這個二次式的陳述。

解答

為簡潔起見，全文記 $x = x_{n+1}$。

第 (i) 部分

按定義 $\sum_{k=1}^{n} x_k = nA$，所以

$$C = \frac{1}{n+1}\left(\sum_{k=1}^{n} x_k + x\right) = \frac{nA + x}{n+1}.$$

第 (ii) 部分

把平方展開，並利用 $\sum_{k=1}^{n} x_k = nA$ 及 $\sum_{k=1}^{n} 1 = n$：

$$\begin{aligned} B &= \frac{1}{n}\sum_{k=1}^{n}(x_k^2 - 2Ax_k + A^2) = \frac{1}{n}\sum_{k=1}^{n} x_k^2 - \frac{2A}{n}\cdot nA + \frac{1}{n}\cdot nA^2 \\ &= \frac{1}{n}\sum_{k=1}^{n} x_k^2 - A^2, \end{aligned}$$

此即所求。

第 (iii) 部分

第 (ii) 部分對任何一組數都成立，所以把它應用於 $n+1$ 個數 $x_1, \ldots, x_{n+1}$（其平均值為 C），得

$$D = \frac{1}{n+1}\sum_{k=1}^{n+1} x_k^2 - C^2.$$

再由第 (ii) 部分，$\sum_{k=1}^{n} x_k^2 = n(B+A^2)$，故 $\sum_{k=1}^{n+1} x_k^2 = n(B+A^2) + x^2$。以第 (i) 部分代入 C，

$$D = \frac{nB + nA^2 + x^2}{n+1} - \frac{(nA+x)^2}{(n+1)^2}.$$

兩邊乘以 $(n+1)^2$ 並展開：

$$\begin{aligned}(n+1)^2 D &= (n+1)(nB + nA^2 + x^2) - (n^2A^2 + 2nAx + x^2)\\ &= n(n+1)B + (n^2 + n - n^2)A^2 - 2nAx + (n+1-1)x^2\\ &= n(n+1)B + n(x^2 - 2Ax + A^2)\\ &= n(n+1)B + n(x-A)^2.\end{aligned}$$

因此

$$\boxed{D = \frac{n}{n+1}B + \frac{n}{(n+1)^2}(x_{n+1} - A)^2.}$$

第一條不等式。 乘以 $n+1$，

$$(n+1)D = nB + \frac{n}{n+1}(x_{n+1} - A)^2 \;\geq\; nB,$$

因為最後一項是一個平方的非負倍數。這對 x_{n+1} 的任何值都成立，而等號成立當且僅當 $x_{n+1} = A$。

? **想一想** 怎樣才能把 $D < B$ 轉化為關於 x_{n+1} 的條件，而又不必擔心負數的平方根？

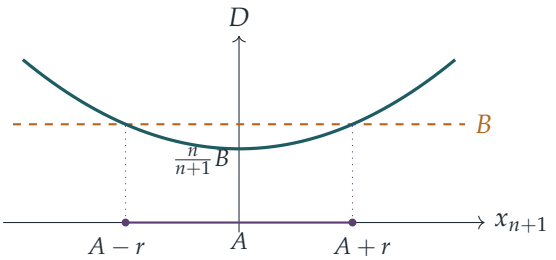
第二條不等式。 首先留意 $B \geq 0$，因為它是一些平方的平均值，所以 $\sqrt{(n+1)B/n}$ 是實數。現在利用方框中的公式，以下每一步都是可逆的（我們只是在兩邊減去同一個量，

以及乘以正數 $(n+1)^2/n$）：

$$\begin{aligned} D < B &\iff \frac{n}{(n+1)^2}(x_{n+1}-A)^2 < B - \frac{n}{n+1}B = \frac{B}{n+1} \\ &\iff (x_{n+1}-A)^2 < \frac{(n+1)B}{n} \\ &\iff |x_{n+1}-A| < \sqrt{\frac{(n+1)B}{n}} \\ &\iff A - \sqrt{\frac{(n+1)B}{n}} < x_{n+1} < A + \sqrt{\frac{(n+1)B}{n}}. \end{aligned}$$

第三步是在兩邊取非負平方根；對非負數之間的嚴格不等式而言，這樣做會保持不等式成立。這就是所求的「當且僅當」。（若 $B = 0$，則該區間是空集，而事實上此時恆有 $D \geq 0 = B$。）

下圖顯示當 $n = 3$、$A = 0$、$B = 1$ 時，D 作為 x_{n+1} 的函數的圖像：這是一條拋物線，在 $x_{n+1} = A$ 處取得最小值 $\frac{n}{n+1}B$，並恰好在上面求得的區間上低於 B 這個水平。



其中 $r = \sqrt{(n+1)B/n}$。

驗算

- 取 $n = 1$：此時 $B = 0$，而公式給出 $D = \frac{1}{4}(x_2 - x_1)^2$；對於兩個數 x_1, x_2 而言這是正確的，因為它們的平均值就是它們的中點。
- 取 $x_1, x_2, x_3 = 0, 0, 3$，則 $A = 1$、$B = 2$，再加入 $x_4 = 1 = A$。直接計算得 $D = \frac{1}{4}(1+1+4+0) = \frac{3}{2} = \frac{3}{4}B$，正是最小值。
- 我們在 20 000 組隨機數列上以分數精確驗證了全部三個結果，包括等價關係 $D < B \iff (x_{n+1}-A)^2 < (n+1)B/n$，並在區間兩端點皆為有理數的 216 個情況中，確認了在兩端點處 $D = B$。

常見錯誤

- 在第 (ii) 部分展開時，忘記 $\sum_{k=1}^{n} A^2 = nA^2$，而不是 A^2。

- 以 $\sum x_k^2$ 表示 D 後便停下來：題目要求以 B、A、x_{n+1} 和 n 表示，所以必須利用第 (ii) 部分把這個和替換掉。
- 對 $(x_{n+1}-A)^2 < k$ 開平方時只得出 $x_{n+1}-A < \sqrt{k}$，漏掉了區間的下端。

STEP 之外

D 的公式是一條**更新公式**（updating formula）：它讓電腦可以逐個數值地維持一個數據流的平均值和方差，而無須儲存全部數據；這個想法在 Welford 算法（Welford's algorithm）中得到進一步改良。$n(x_{n+1}-A)^2/(n+1)^2$ 這一項，是把方差分解為「組內」和「組間」兩部分的一個小例子，而這種分解正是統計學中方差分析（analysis of variance）的基礎。如欲深入了解，可參閱 Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率導論》）。

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。