详细解答

STEP III 2010，第 1 题

纯数学 · 平均值与方差

> **STEP III 2010，第 1 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/10-s3-q1

关键思路

A 和 B 是首 n 个数的平均值（mean）和方差（variance）；C 和 D 则是再加入一个数之后的平均值和方差。方差最容易以“平方的平均值减去平均值的平方”这个形式处理，而这正是第 (ii) 部分给我们的结果。有了这个形式，D 便成为 x_{n+1} 的二次式（quadratic），而两条不等式（inequality）都是关于这个二次式的陈述。

解答

为简洁起见，全文记 $x = x_{n+1}$。

第 (i) 部分

按定义 $\sum_{k=1}^{n} x_k = nA$，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

把平方展开，并利用 $\sum_{k=1}^{n} x_k = nA$ 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) 部分对任何一组数都成立，所以把它应用于 $n+1$ 个数 $x_1,\ldots,x_{n+1}$（其平均值为 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，

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

两边乘以 $(n+1)^2$ 并展开：

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

因此

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

第一条不等式。 乘以 $n+1$，

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

因为最后一项是一个平方的非负倍数。这对 x_{n+1} 的任何值都成立，而等号成立当且仅当 $x_{n+1}=A$。

? **想一想** 怎样才能把 $D<B$ 转化为关于 x_{n+1} 的条件，而又不必担心负数的平方根？

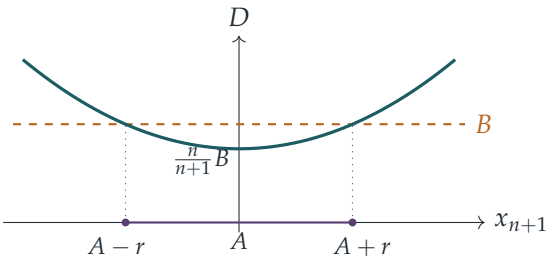
第二条不等式。 首先留意 $B\geq 0$，因为它是一些平方的平均值，所以 $\sqrt{(n+1)B/n}$ 是实数。现在利用方框中的公式，以下每一步都是可逆的（我们只是在两边减去同一个量，以

及乘以正数 $(n+1)^2/n$）：

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

第三步是在两边取非负平方根；对非负数之间的严格不等式而言，这样做会保持不等式成立。这就是所求的“当且仅当”。（若 $B = 0$，则该区间是空集，而事实上此时恒有 $D \geq 0 = B$。）

下图显示当 $n = 3$、$A = 0$、$B = 1$ 时，D 作为 x_{n+1} 的函数的图像：这是一条抛物线，在 $x_{n+1} = A$ 处取得最小值 $\frac{n}{n+1}B$，并恰好在上面求得的区间上低于 B 这个水平。



其中 $r = \sqrt{(n+1)B/n}$。

验算

- 取 $n = 1$：此时 $B = 0$，而公式给出 $D = \frac{1}{4}(x_2 - x_1)^2$；对于两个数 x_1, x_2 而言这是正确的，因为它们的平均值就是它们的中点。
- 取 $x_1, x_2, x_3 = 0, 0, 3$，则 $A = 1$、$B = 2$，再加入 $x_4 = 1 = A$。直接计算得 $D = \frac{1}{4}(1+1+4+0) = \frac{3}{2} = \frac{3}{4}B$，正是最小值。
- 我们在 20 000 组随机数列上以分数精确验证了全部三个结果，包括等价关系 $D < B \iff (x_{n+1}-A)^2 < (n+1)B/n$，并在区间两端点皆为有理数的 216 个情况中，确认了在两端点处 $D = B$。

常见错误

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

- 以 $\sum x_k^2$ 表示 D 后便停下来：题目要求以 B、A、x_{n+1} 和 n 表示，所以必须利用第 (ii) 部分把这个和替换掉。
- 对 $(x_{n+1}-A)^2 < k$ 开平方时只得出 $x_{n+1}-A < \sqrt{k}$，漏掉了区间的下端。

STEP 之外

D 的公式是一条**更新公式**（updating formula）：它让电脑可以逐个数值地维持一个数据流的平均值和方差，而无须储存全部数据；这个想法在 Welford 算法（Welford's algorithm）中得到进一步改良。$n(x_{n+1}-A)^2/(n+1)^2$ 这一项，是把方差分解为“组内”和“组间”两部分的一个小例子，而这种分解正是统计学中方差分析（analysis of variance）的基础。如欲深入了解，可参阅 Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率导论》）。

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。